(12) UK Patent Application (19) GB (11) 2 349 762 (13) A

(43) Date of A Publication 08.11.2000

(21) Application No 9905191.4

(22) Date of Filing 05.03.1999

(51) INT CL[7]
G06T 15/00

(52) UK CL (Edition R)
H4F FDX F12S F15 F22 F3P F3T F3X F30J F31A F32

(56) Documents Cited
WO 99/65223 A2 US 5819286 A US 5499146 A
http://garuda.imag.fr/MPEG4/syssite/syspub/version1/index

(58) Field of Search
UK CL (Edition R) G4A AUDB , H4F FAAX FDX FGXD FGXX FRG FRX
INT CL[7] G06T 15/00
ONLINE: WPI; JAPIO; EPODOC; INSPEC

(71) Applicant(s)
Canon Kabushiki Kaisha
(Incorporated in Japan)
30-2 3-chome Shimomaruko, Ohta-ku, Tokyo, Japan

(72) Inventor(s)
Simon Michael Rowe
Michael James Taylor

(74) Agent and/or Address for Service
Beresford & Co
2-5 Warwick Court, High Holborn, LONDON,
WC1R 5DJ, United Kingdom

(54) Abstract Title
3-D image archiving apparatus

(57) In an image processing apparatus 32, data defining the three-dimensional movements of people is processed to animate three-dimensional computer models of those people. Data defining the time-evolving three-dimensional computer models and sound data received from the real-world people is stored in an archive database (preferably in MPEG4 format) 126 as data 502, together with additional information to facilitate the searching and retrieval of data therefrom. The additional information includes text 504 which comprises the words spoken by each real-world person and is produced using voice recognition processing, view parameters 512 which define the person or object at which each real-world person is looking during each predetermined period of time, and viewing histograms 540 which, for each period of text 504 define the percentage of time that the speaking person spent looking at each other real-world person. The storage of the text 504, view parameters 512 and viewing histograms 540 facilitates the improved searching and retrieval of data 502 from the archive database 126.

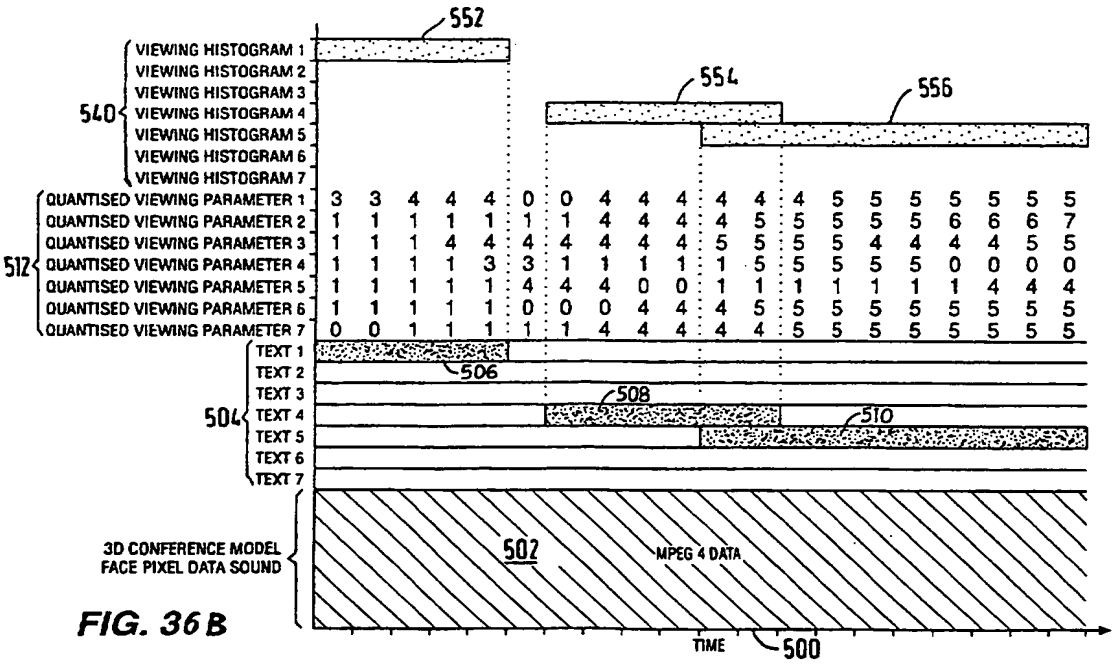


FIG. 36B

GB 2 349 762 A

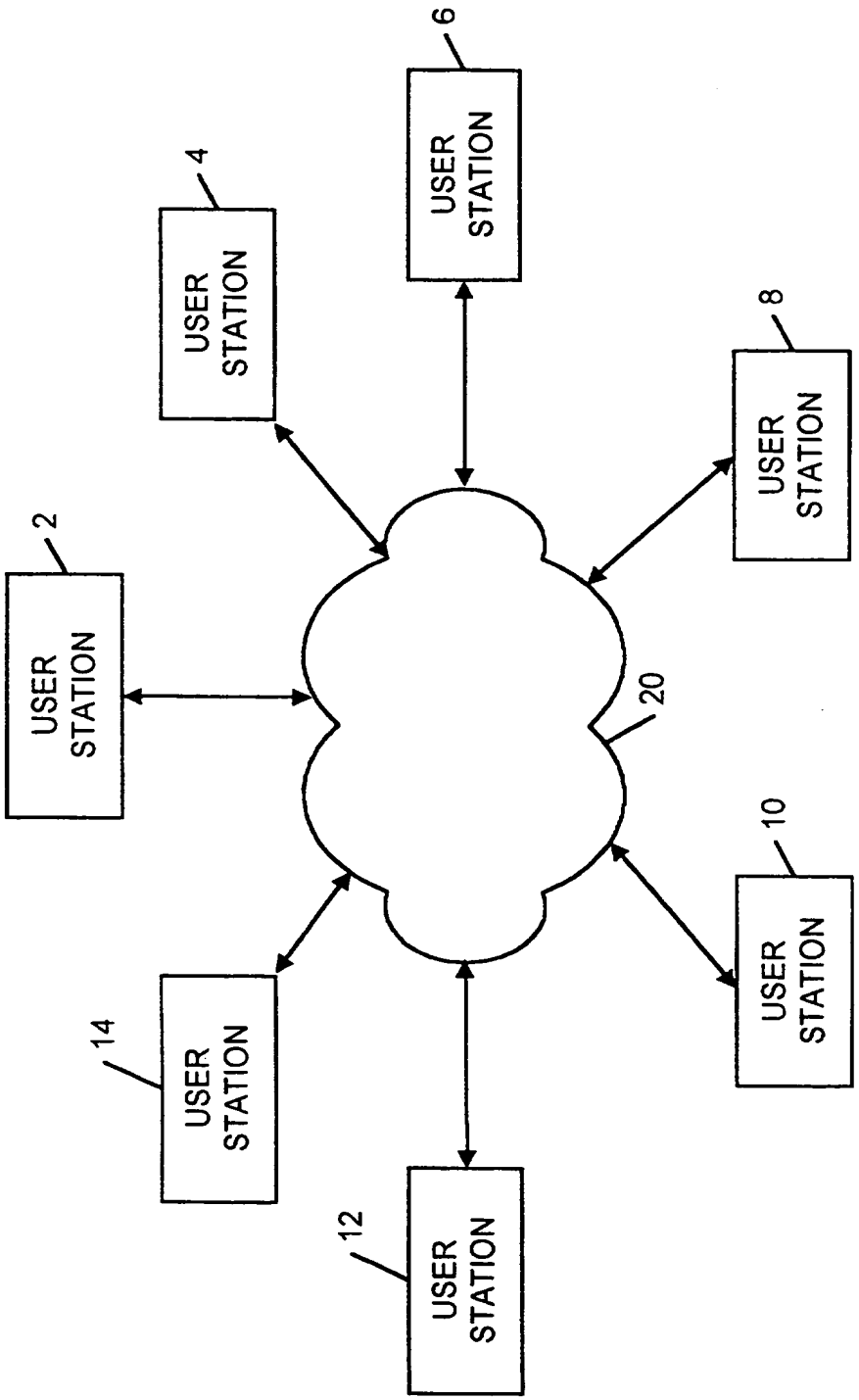

USER STATION
2
USER STATION
4
USER STATION
6
USER STATION
8
USER STATION
10
USER STATION
12
USER STATION
14
20

FIG. 1

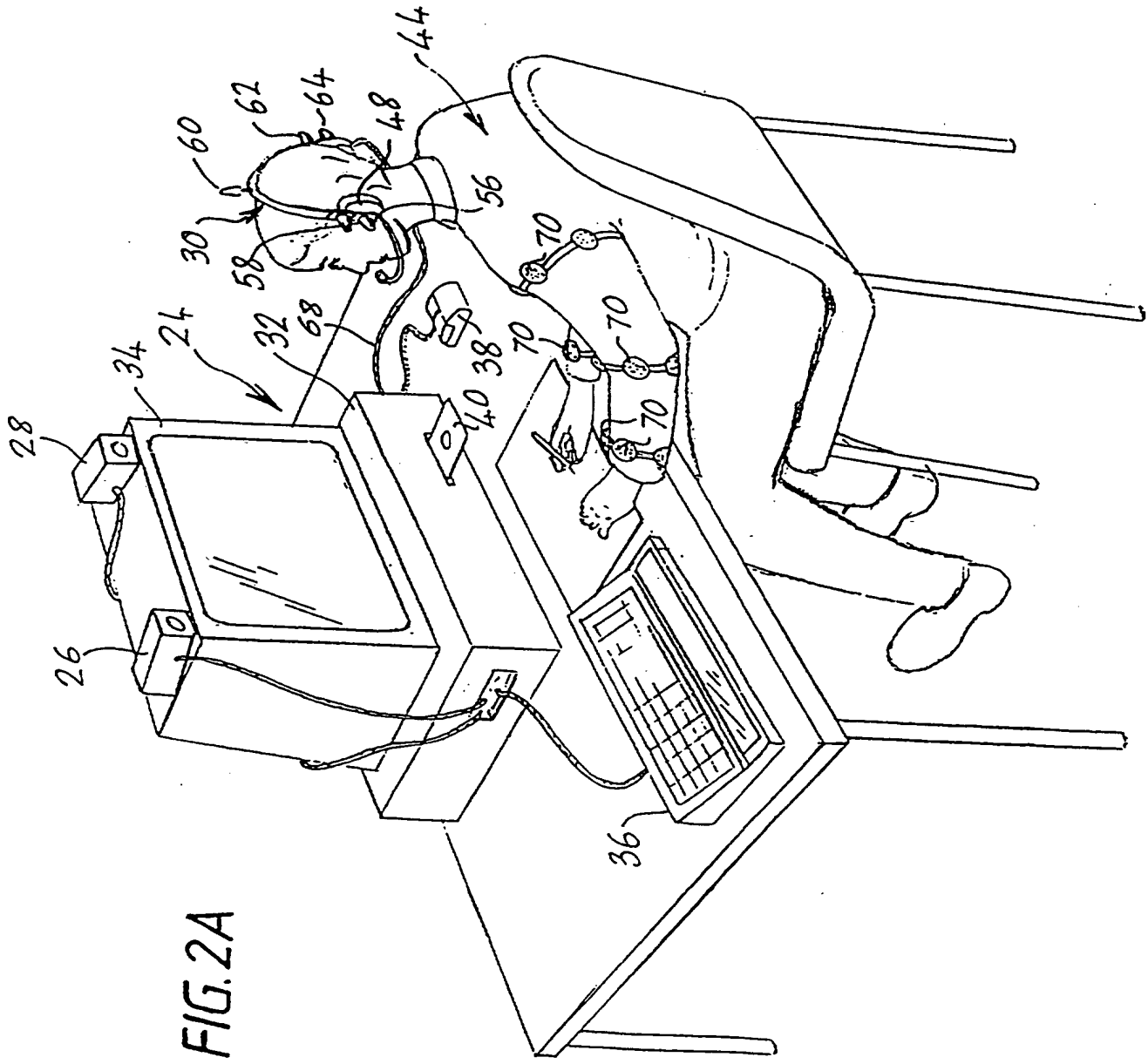


FIG. 2A

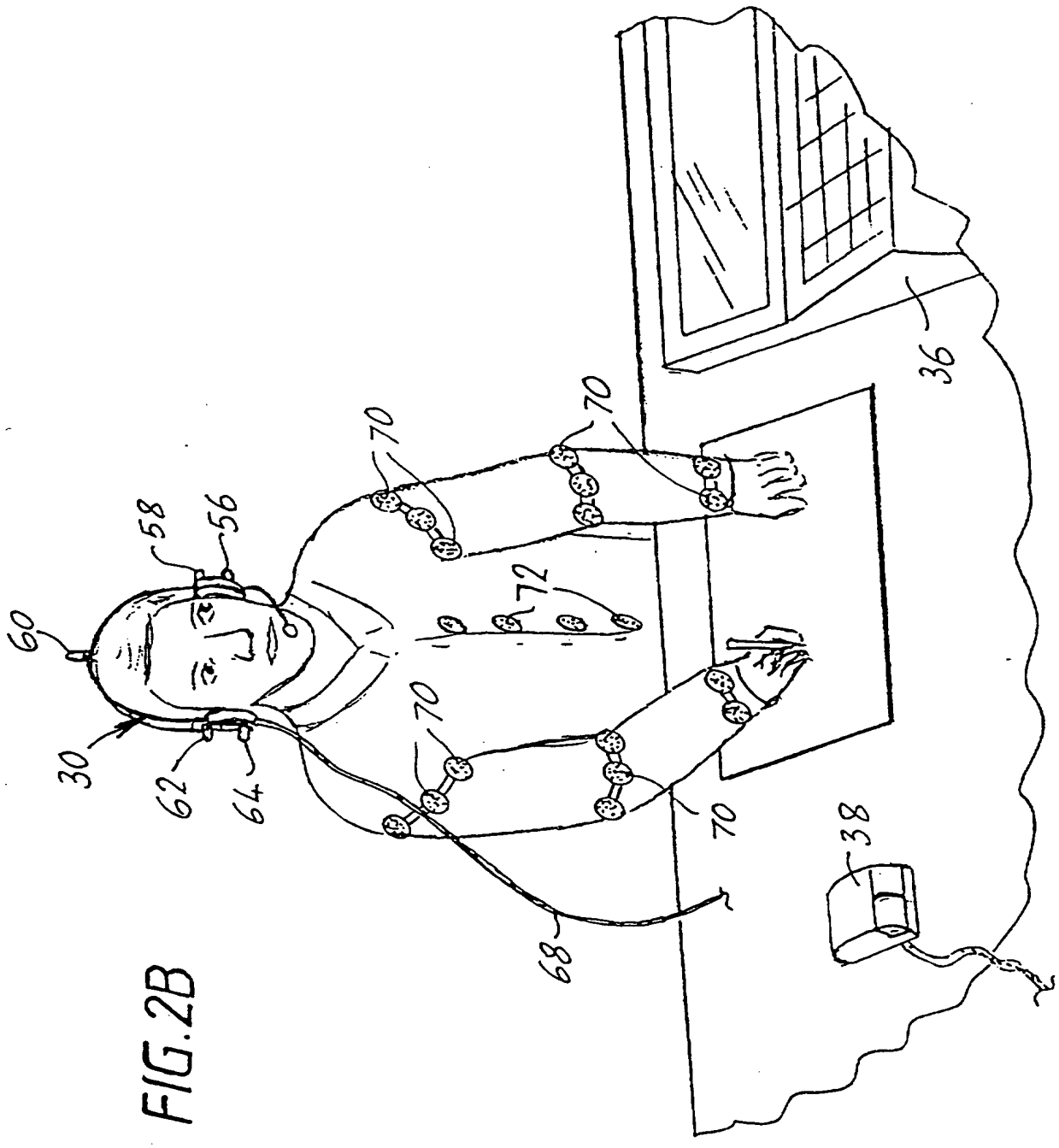


FIG. 2B

FIG. 2C

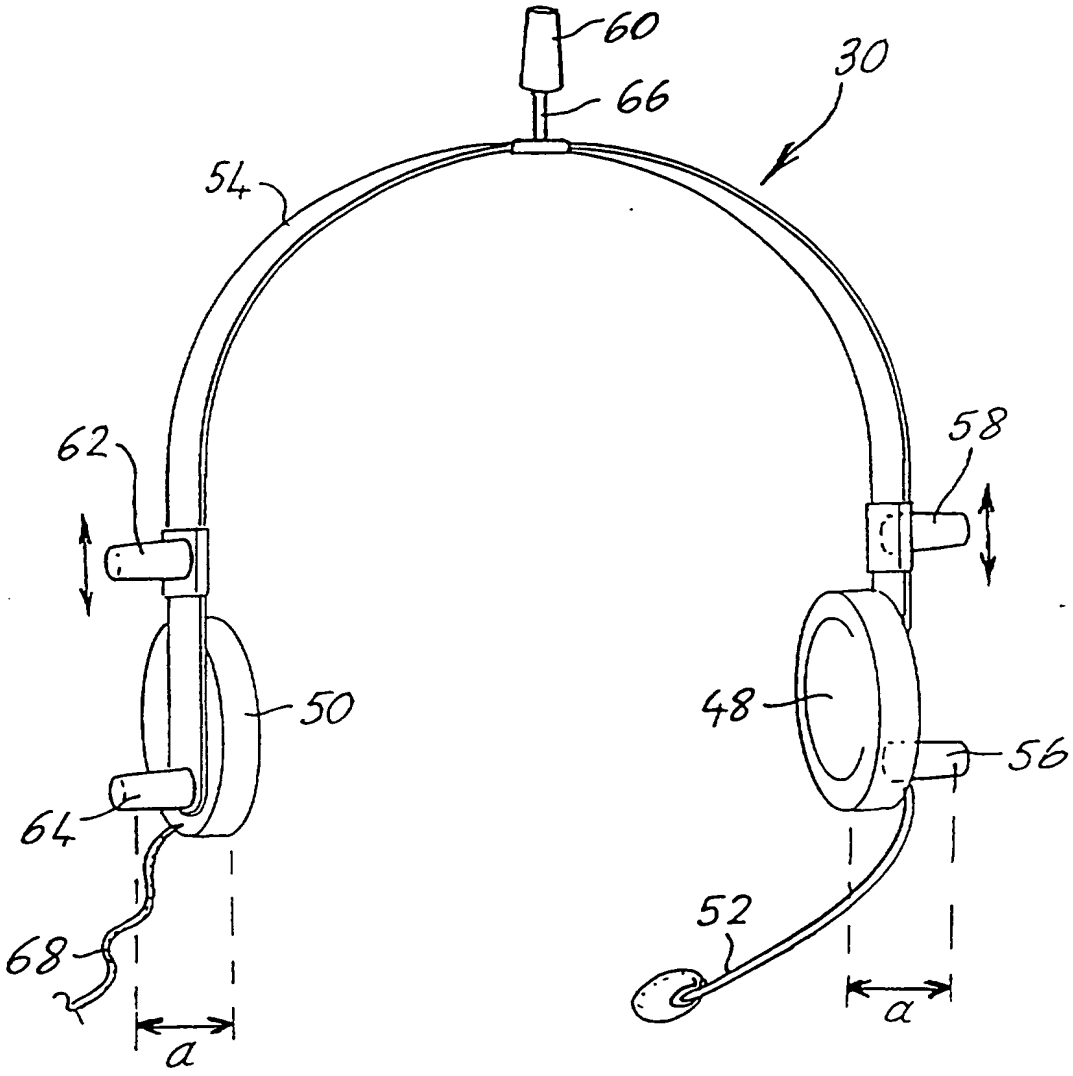
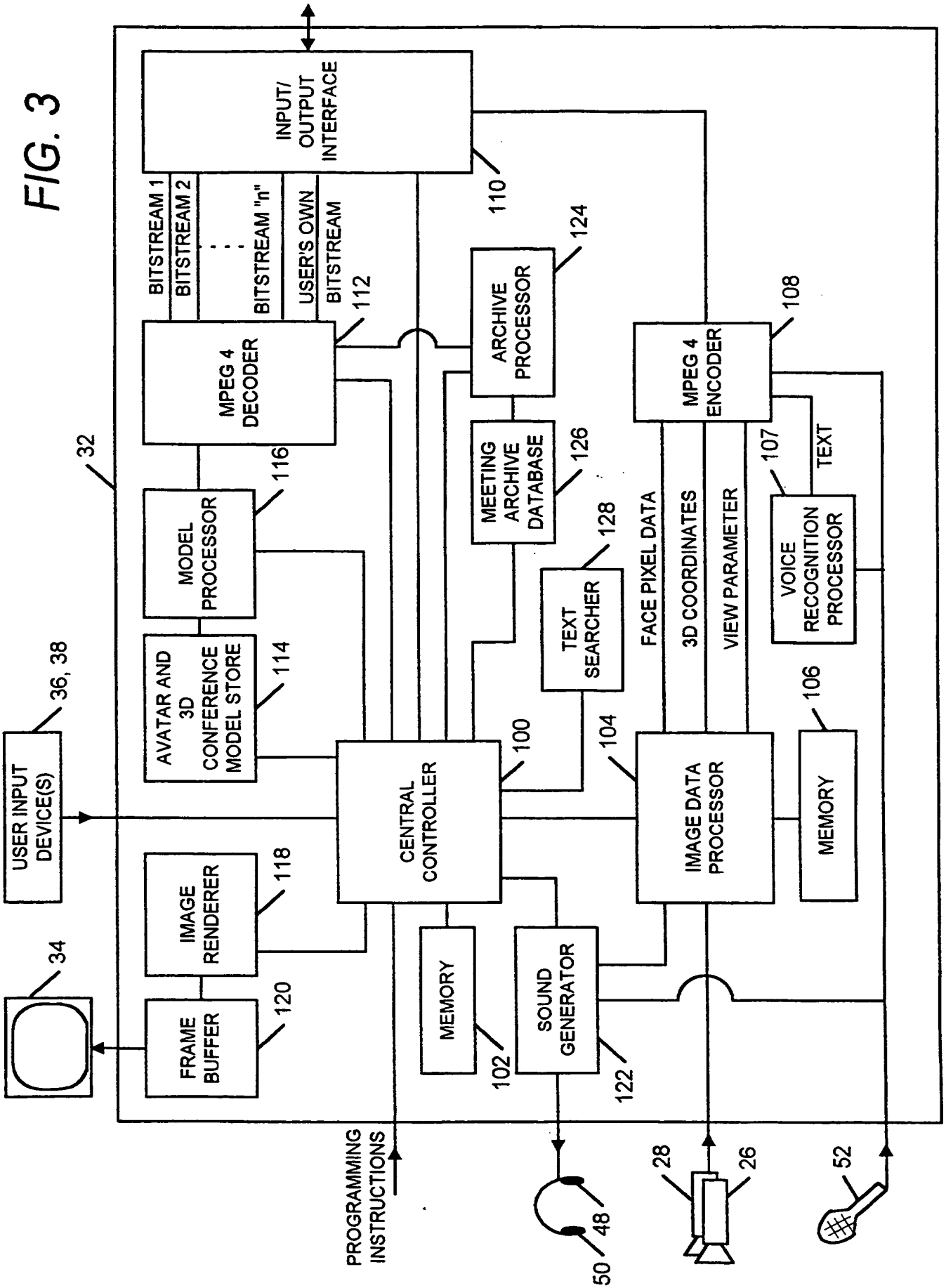


FIG. 3


INPUT/ OUTPUT INTERFACE
110
BITSTREAM 1
BITSTREAM 2
BITSTREAM "n"
USER'S OWN BITSTREAM
112
MPEG 4 DECODER
116
MODEL PROCESSOR
AVATAR AND 3D CONFERENCE MODEL STORE
114
ARCHIVE PROCESSOR
124
MEETING ARCHIVE DATABASE
126
TEXT SEARCHER
128
MPEG 4 ENCODER
108
FACE PIXEL DATA
3D COORDINATES
VIEW PARAMETER
107
TEXT
VOICE RECOGNITION PROCESSOR
32
36, 38
USER INPUT DEVICE(S)
CENTRAL CONTROLLER
100
104
IMAGE DATA PROCESSOR
106
MEMORY
IMAGE RENDERER
118
34
FRAME BUFFER
120
MEMORY
102
SOUND GENERATOR
122
PROGRAMMING INSTRUCTIONS
48
50
28
26
52

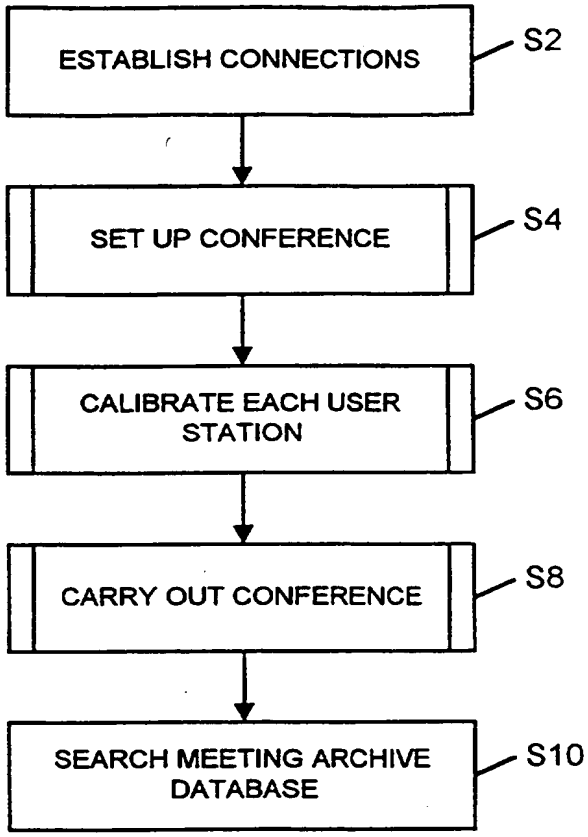

ESTABLISH CONNECTIONS
S2
SET UP CONFERENCE
S4
CALIBRATE EACH USER STATION
S6
CARRY OUT CONFERENCE
S8
SEARCH MEETING ARCHIVE DATABASE
S10

FIG. 4

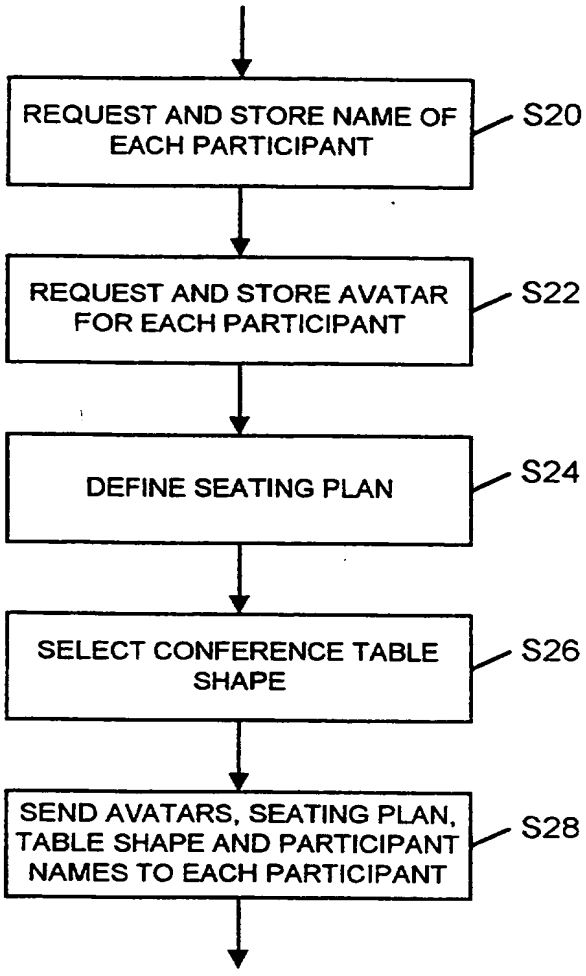

REQUEST AND STORE NAME OF EACH PARTICIPANT
S20
REQUEST AND STORE AVATAR FOR EACH PARTICIPANT
S22
DEFINE SEATING PLAN
S24
SELECT CONFERENCE TABLE SHAPE
S26
SEND AVATARS, SEATING PLAN, TABLE SHAPE AND PARTICIPANT NAMES TO EACH PARTICIPANT
S28

FIG. 5

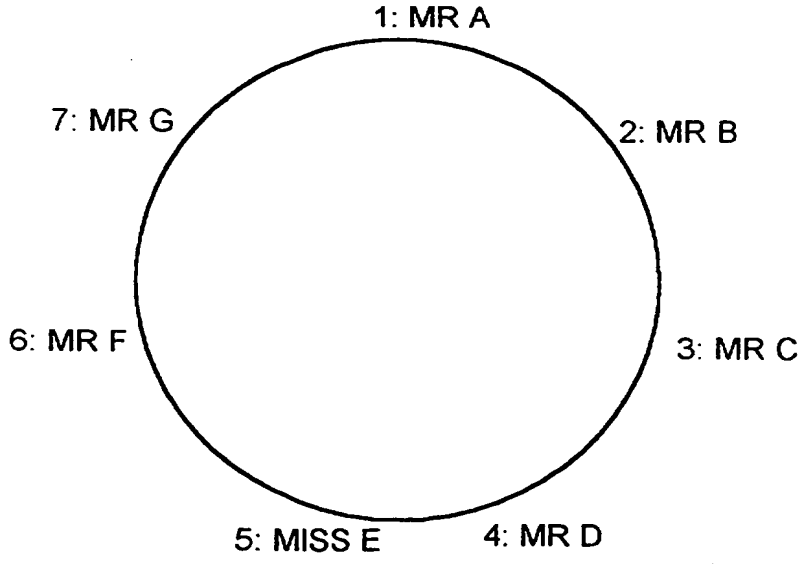

1: MR A
2: MR B
3: MR C
4: MR D
5: MISS E
6: MR F
7: MR G

FIG. 6

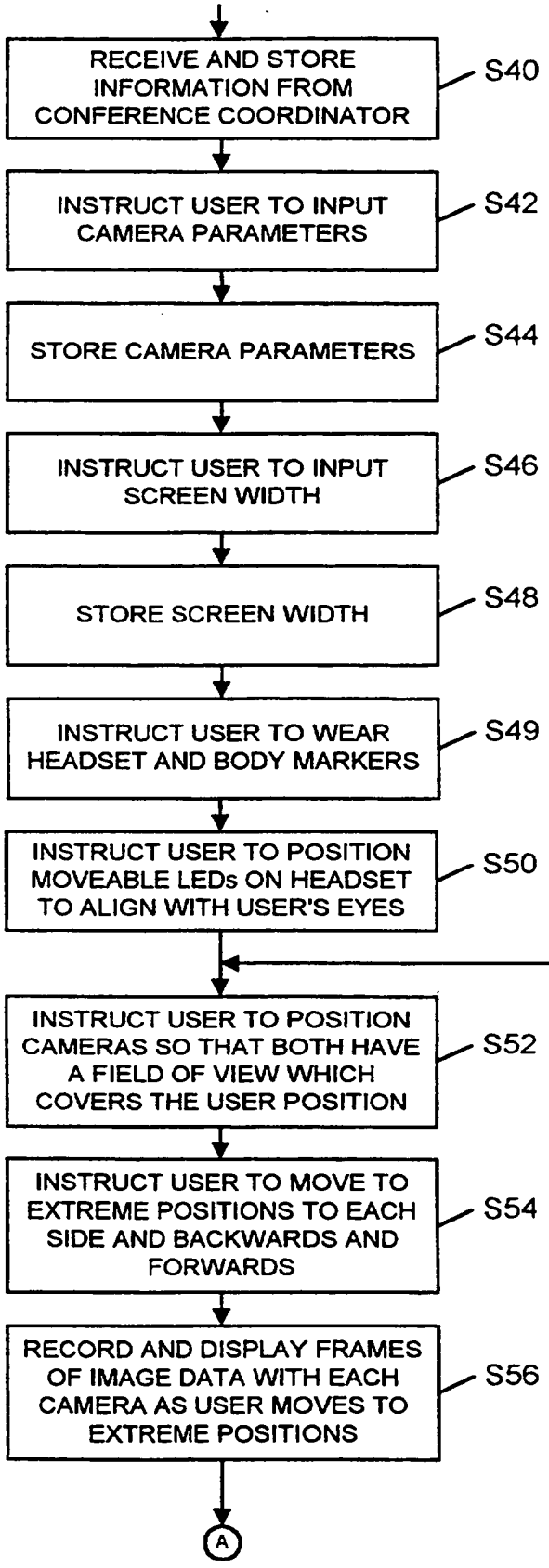

RECEIVE AND STORE INFORMATION FROM CONFERENCE COORDINATOR
S40
INSTRUCT USER TO INPUT CAMERA PARAMETERS
S42
STORE CAMERA PARAMETERS
S44
INSTRUCT USER TO INPUT SCREEN WIDTH
S46
STORE SCREEN WIDTH
S48
INSTRUCT USER TO WEAR HEADSET AND BODY MARKERS
S49
INSTRUCT USER TO POSITION MOVEABLE LEDs ON HEADSET TO ALIGN WITH USER'S EYES
S50
INSTRUCT USER TO POSITION CAMERAS SO THAT BOTH HAVE A FIELD OF VIEW WHICH COVERS THE USER POSITION
S52
INSTRUCT USER TO MOVE TO EXTREME POSITIONS TO EACH SIDE AND BACKWARDS AND FORWARDS
S54
RECORD AND DISPLAY FRAMES OF IMAGE DATA WITH EACH CAMERA AS USER MOVES TO EXTREME POSITIONS
S56
A
B

FIG. 7

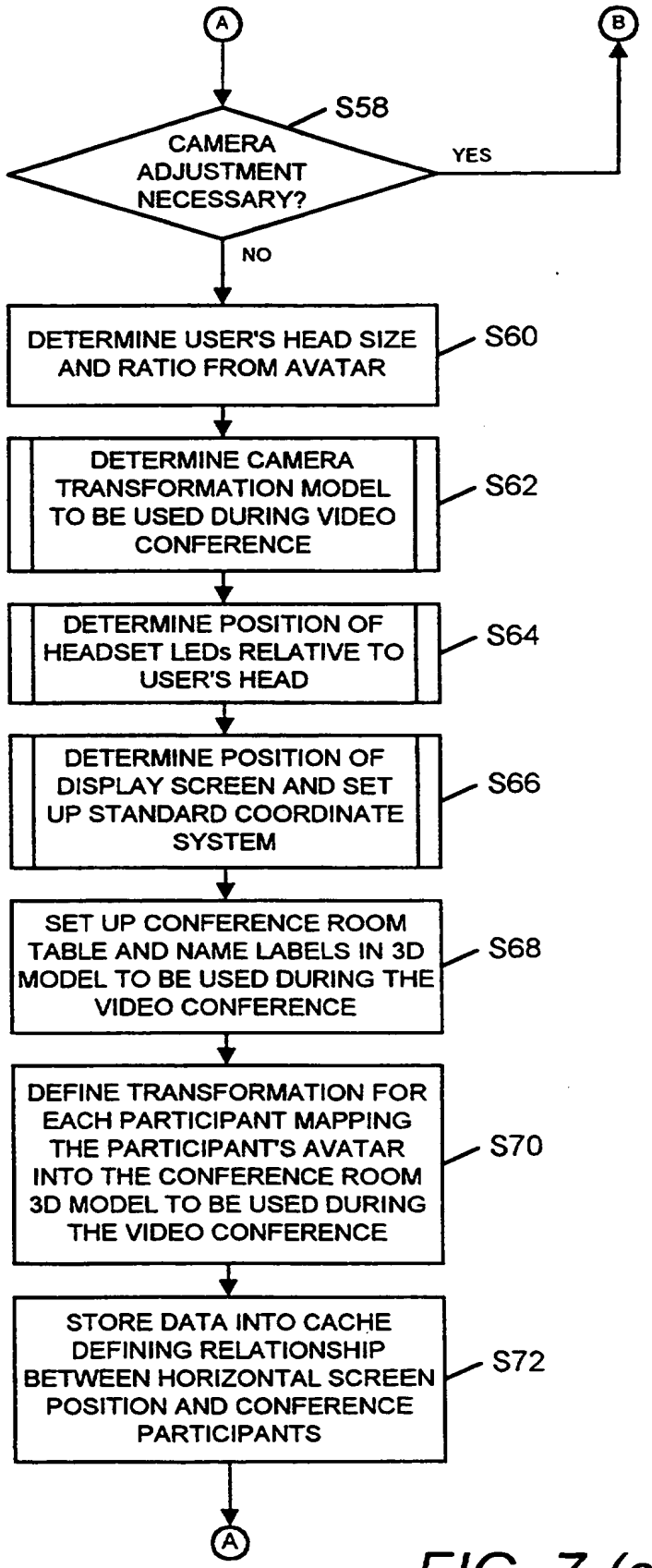

A
B
S58
CAMERA ADJUSTMENT NECESSARY?
YES
NO
DETERMINE USER'S HEAD SIZE AND RATIO FROM AVATAR
S60
DETERMINE CAMERA TRANSFORMATION MODEL TO BE USED DURING VIDEO CONFERENCE
S62
DETERMINE POSITION OF HEADSET LEDs RELATIVE TO USER'S HEAD
S64
DETERMINE POSITION OF DISPLAY SCREEN AND SET UP STANDARD COORDINATE SYSTEM
S66
SET UP CONFERENCE ROOM TABLE AND NAME LABELS IN 3D MODEL TO BE USED DURING THE VIDEO CONFERENCE
S68
DEFINE TRANSFORMATION FOR EACH PARTICIPANT MAPPING THE PARTICIPANT'S AVATAR INTO THE CONFERENCE ROOM 3D MODEL TO BE USED DURING THE VIDEO CONFERENCE
S70
STORE DATA INTO CACHE DEFINING RELATIONSHIP BETWEEN HORIZONTAL SCREEN POSITION AND CONFERENCE PARTICIPANTS
S72
A

FIG. 7 (cont)

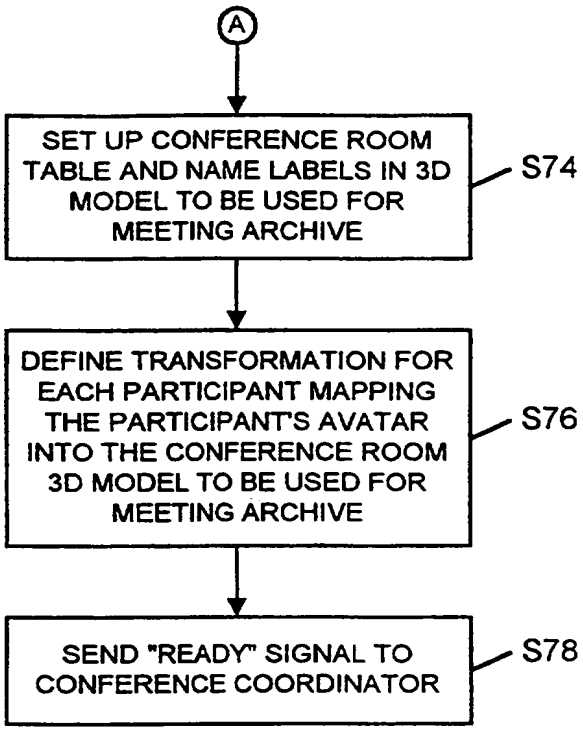

A
SET UP CONFERENCE ROOM TABLE AND NAME LABELS IN 3D MODEL TO BE USED FOR MEETING ARCHIVE
S74
DEFINE TRANSFORMATION FOR EACH PARTICIPANT MAPPING THE PARTICIPANT'S AVATAR INTO THE CONFERENCE ROOM 3D MODEL TO BE USED FOR MEETING ARCHIVE
S76
SEND "READY" SIGNAL TO CONFERENCE COORDINATOR
S78

FIG. 7 (cont)

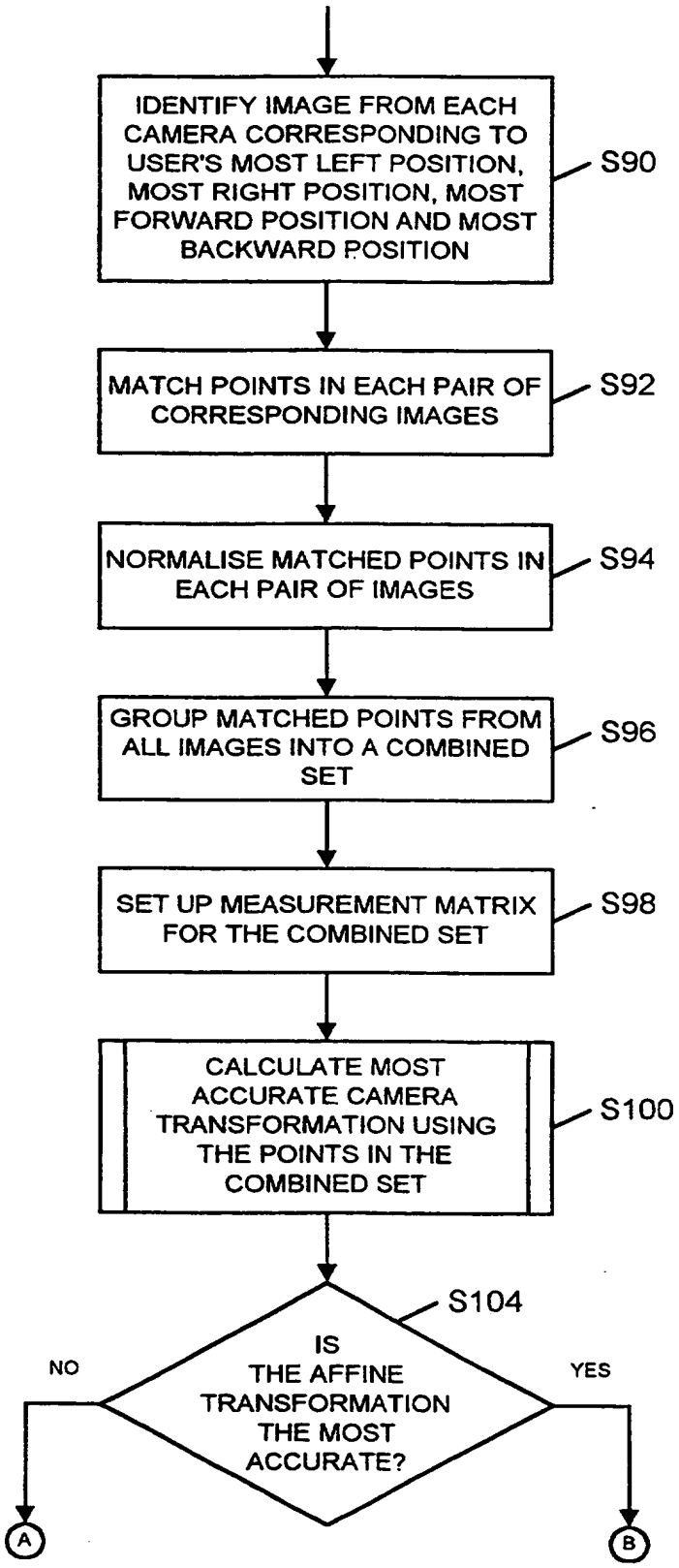

IDENTIFY IMAGE FROM EACH CAMERA CORRESPONDING TO USER'S MOST LEFT POSITION, MOST RIGHT POSITION, MOST FORWARD POSITION AND MOST BACKWARD POSITION
S90
MATCH POINTS IN EACH PAIR OF CORRESPONDING IMAGES
S92
NORMALISE MATCHED POINTS IN EACH PAIR OF IMAGES
S94
GROUP MATCHED POINTS FROM ALL IMAGES INTO A COMBINED SET
S96
SET UP MEASUREMENT MATRIX FOR THE COMBINED SET
S98
CALCULATE MOST ACCURATE CAMERA TRANSFORMATION USING THE POINTS IN THE COMBINED SET
S100
S104
IS THE AFFINE TRANSFORMATION THE MOST ACCURATE?
NO
YES
A
B

FIG. 8

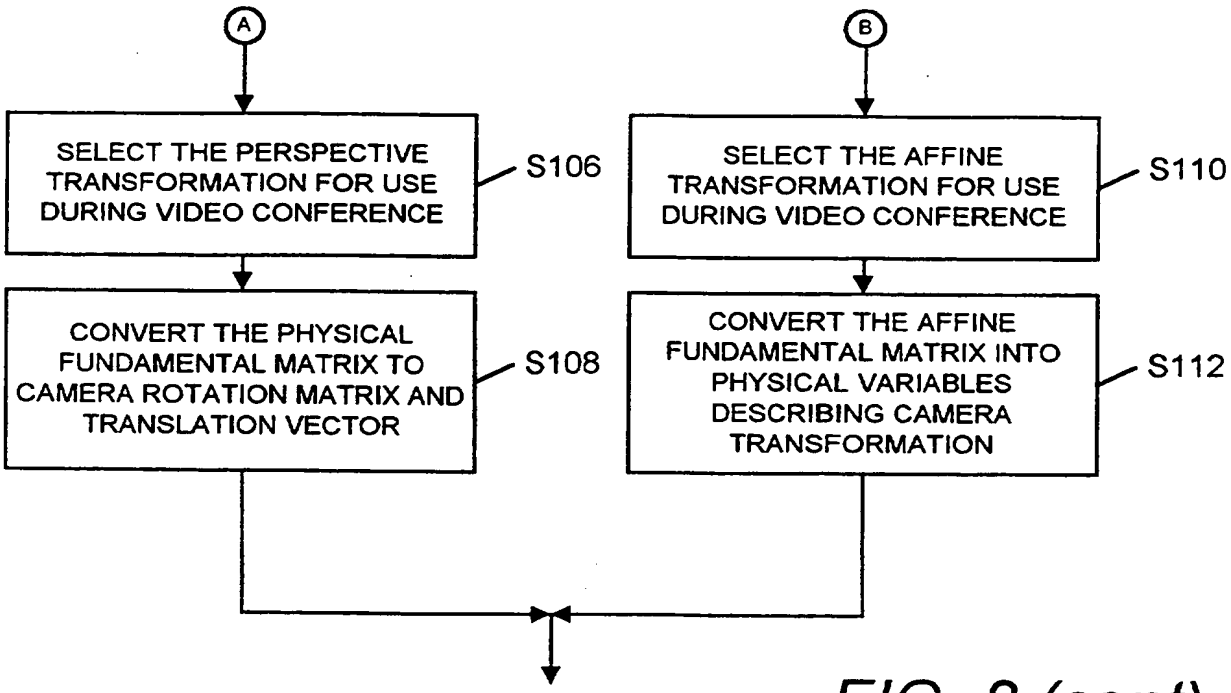


FIG. 8 (cont)

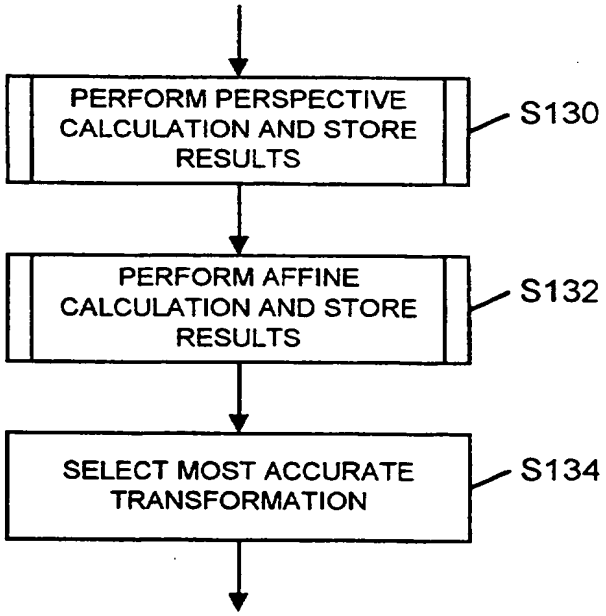


FIG. 9

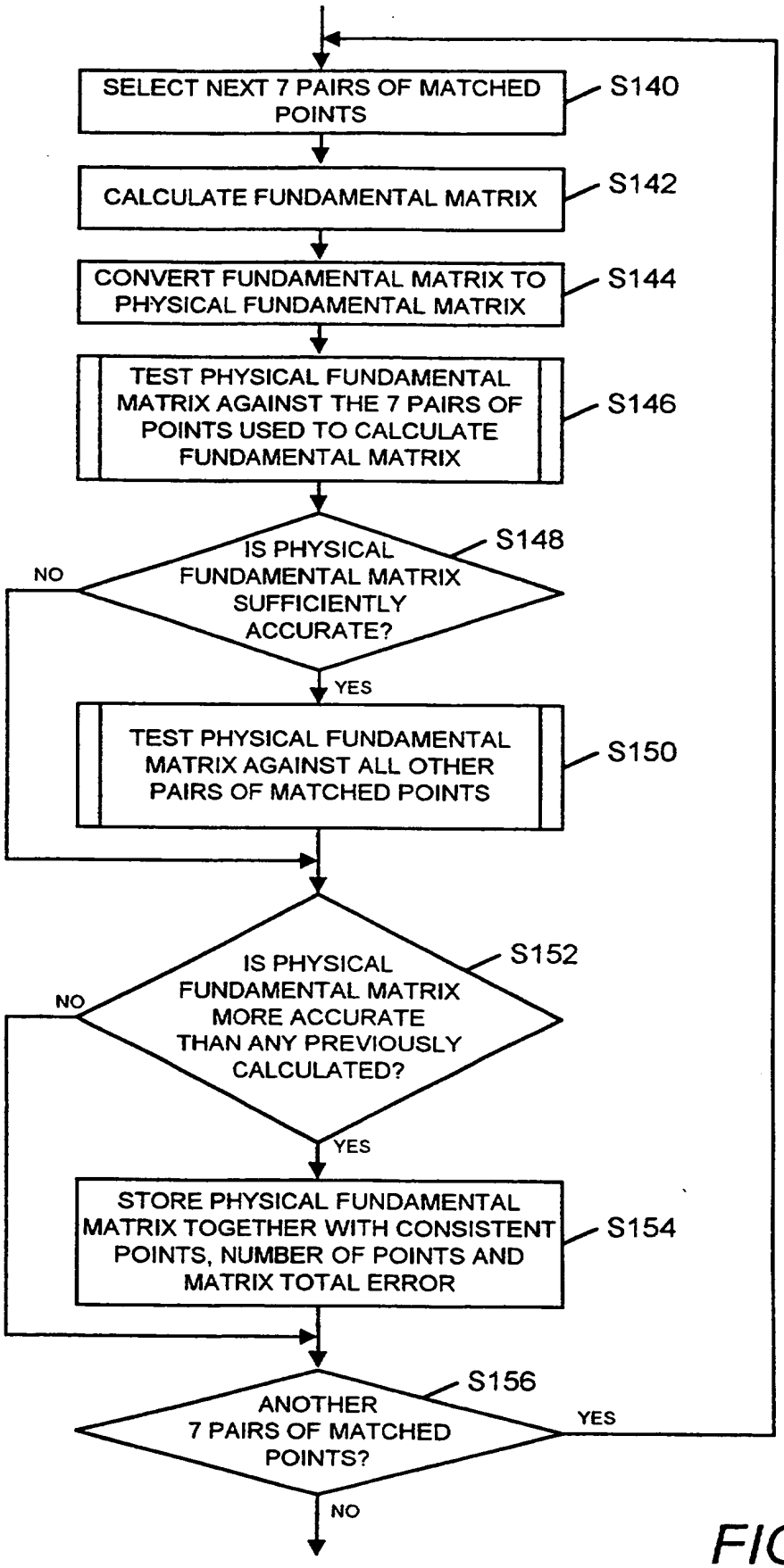

SELECT NEXT 7 PAIRS OF MATCHED POINTS
S140
CALCULATE FUNDAMENTAL MATRIX
S142
CONVERT FUNDAMENTAL MATRIX TO PHYSICAL FUNDAMENTAL MATRIX
S144
TEST PHYSICAL FUNDAMENTAL MATRIX AGAINST THE 7 PAIRS OF POINTS USED TO CALCULATE FUNDAMENTAL MATRIX
S146
IS PHYSICAL FUNDAMENTAL MATRIX SUFFICIENTLY ACCURATE?
S148
NO
YES
TEST PHYSICAL FUNDAMENTAL MATRIX AGAINST ALL OTHER PAIRS OF MATCHED POINTS
S150
IS PHYSICAL FUNDAMENTAL MATRIX MORE ACCURATE THAN ANY PREVIOUSLY CALCULATED?
S152
NO
YES
STORE PHYSICAL FUNDAMENTAL MATRIX TOGETHER WITH CONSISTENT POINTS, NUMBER OF POINTS AND MATRIX TOTAL ERROR
S154
ANOTHER 7 PAIRS OF MATCHED POINTS?
S156
YES
NO

FIG. 10

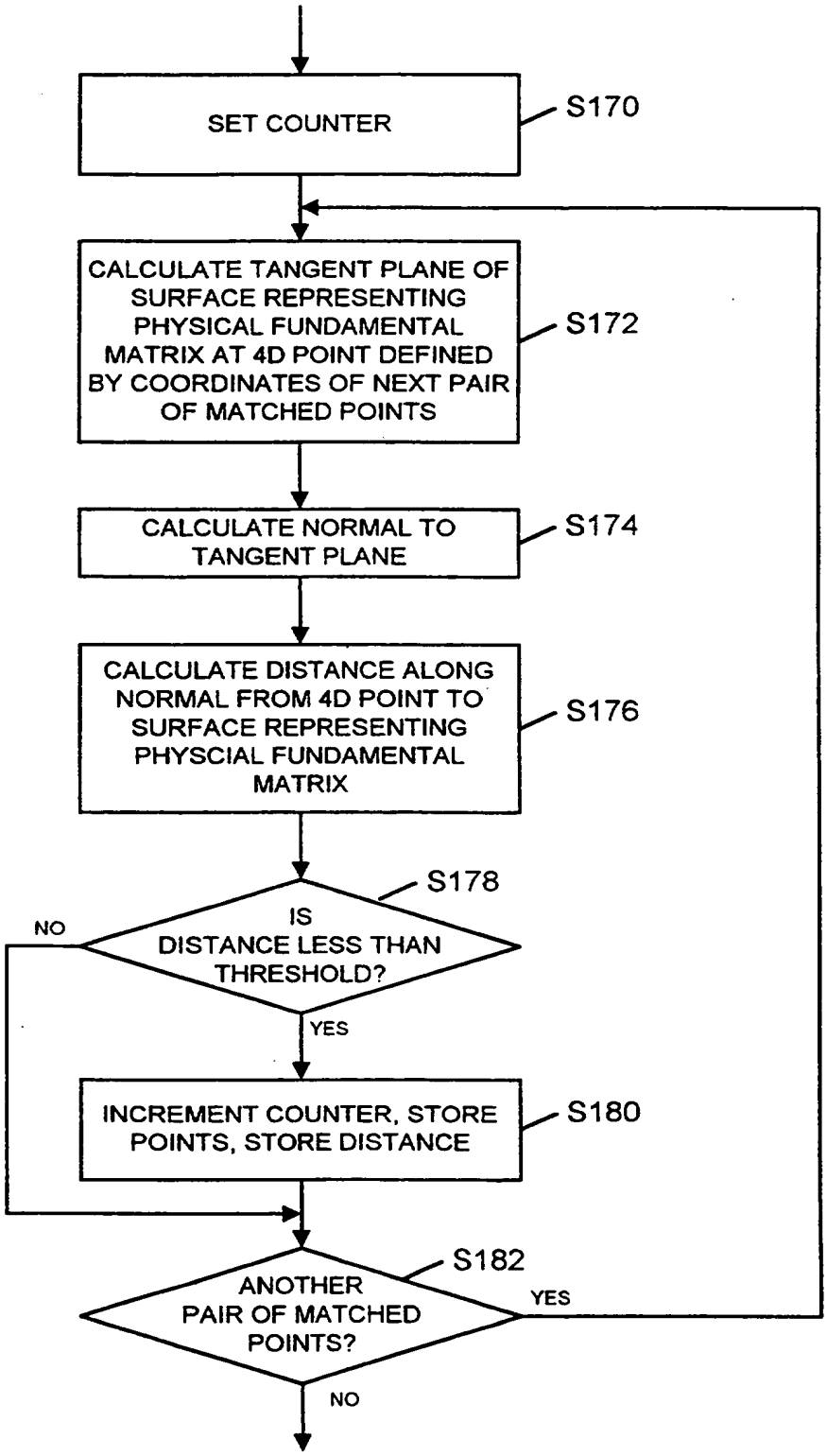

SET COUNTER
S170
CALCULATE TANGENT PLANE OF SURFACE REPRESENTING PHYSICAL FUNDAMENTAL MATRIX AT 4D POINT DEFINED BY COORDINATES OF NEXT PAIR OF MATCHED POINTS
S172
CALCULATE NORMAL TO TANGENT PLANE
S174
CALCULATE DISTANCE ALONG NORMAL FROM 4D POINT TO SURFACE REPRESENTING PHYSCIAL FUNDAMENTAL MATRIX
S176
S178
IS DISTANCE LESS THAN THRESHOLD?
NO
YES
INCREMENT COUNTER, STORE POINTS, STORE DISTANCE
S180
S182
ANOTHER PAIR OF MATCHED POINTS?
YES
NO

FIG. 11

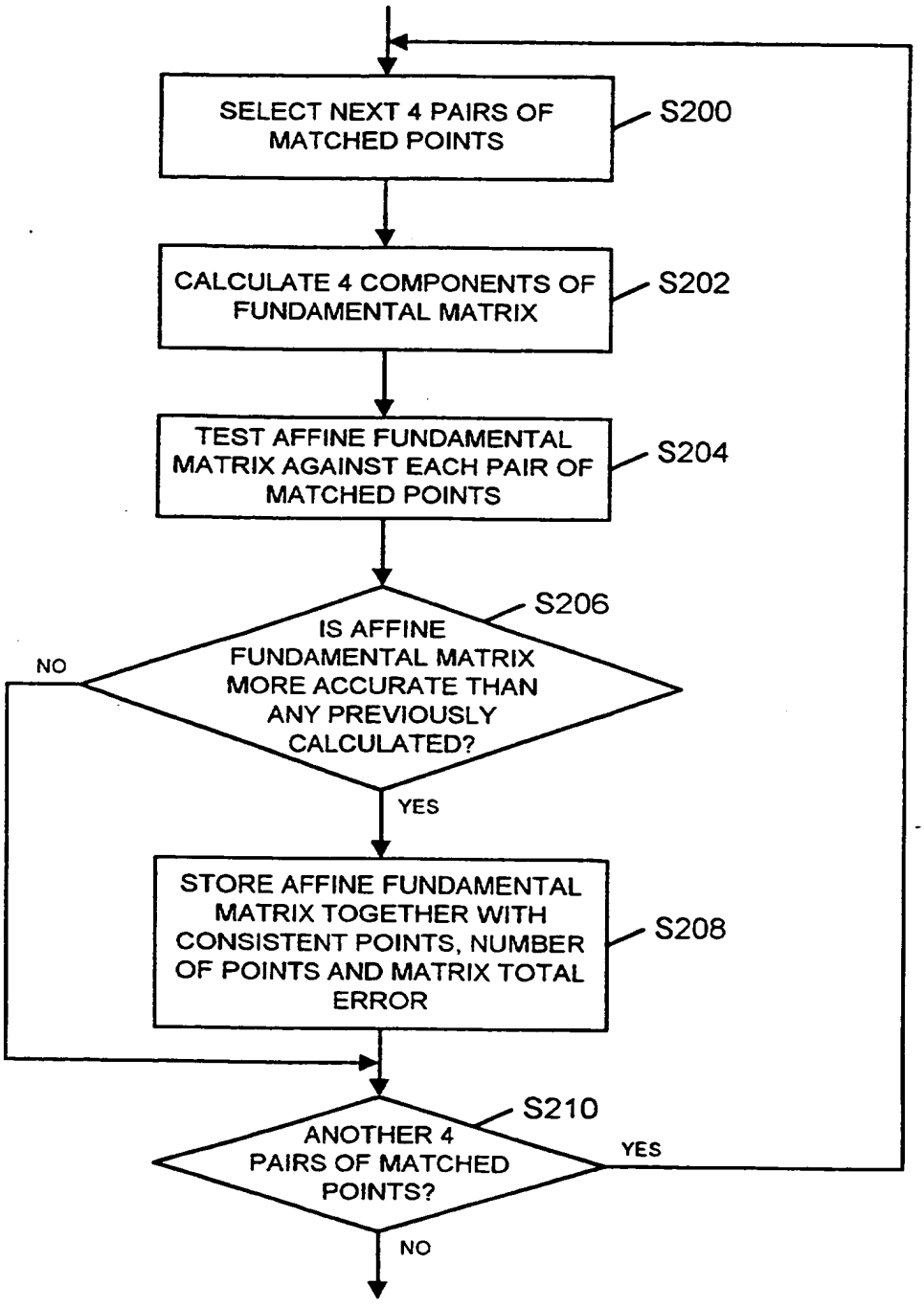

SELECT NEXT 4 PAIRS OF MATCHED POINTS
S200
CALCULATE 4 COMPONENTS OF FUNDAMENTAL MATRIX
S202
TEST AFFINE FUNDAMENTAL MATRIX AGAINST EACH PAIR OF MATCHED POINTS
S204
IS AFFINE FUNDAMENTAL MATRIX MORE ACCURATE THAN ANY PREVIOUSLY CALCULATED?
S206
NO
YES
STORE AFFINE FUNDAMENTAL MATRIX TOGETHER WITH CONSISTENT POINTS, NUMBER OF POINTS AND MATRIX TOTAL ERROR
S208
ANOTHER 4 PAIRS OF MATCHED POINTS?
S210
YES
NO

FIG. 12

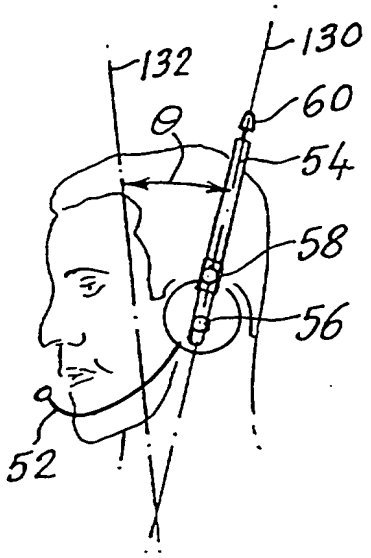

132
130
60
θ
54
58
56
52

FIG. 13

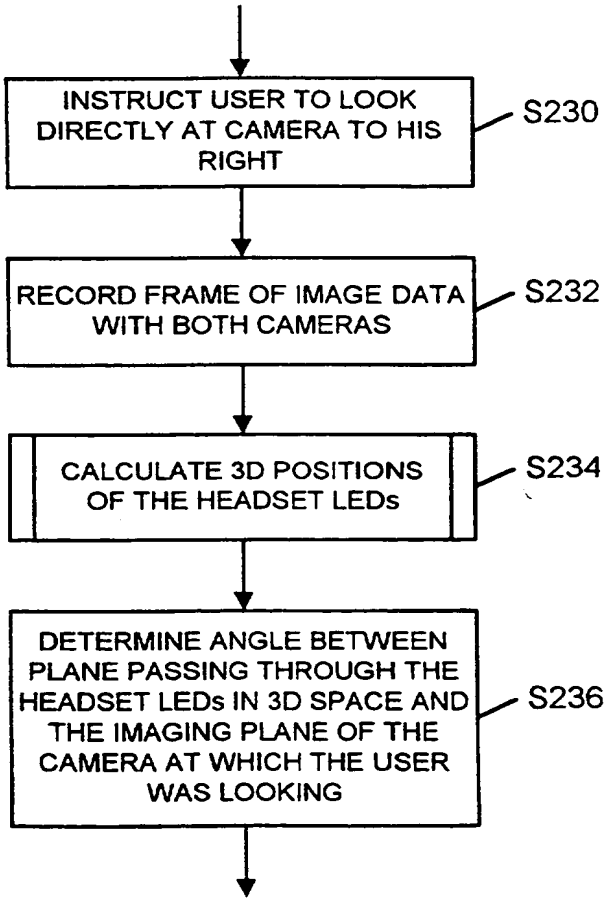

INSTRUCT USER TO LOOK DIRECTLY AT CAMERA TO HIS RIGHT
S230
RECORD FRAME OF IMAGE DATA WITH BOTH CAMERAS
S232
CALCULATE 3D POSITIONS OF THE HEADSET LEDs
S234
DETERMINE ANGLE BETWEEN PLANE PASSING THROUGH THE HEADSET LEDs IN 3D SPACE AND THE IMAGING PLANE OF THE CAMERA AT WHICH THE USER WAS LOOKING
S236

FIG. 14

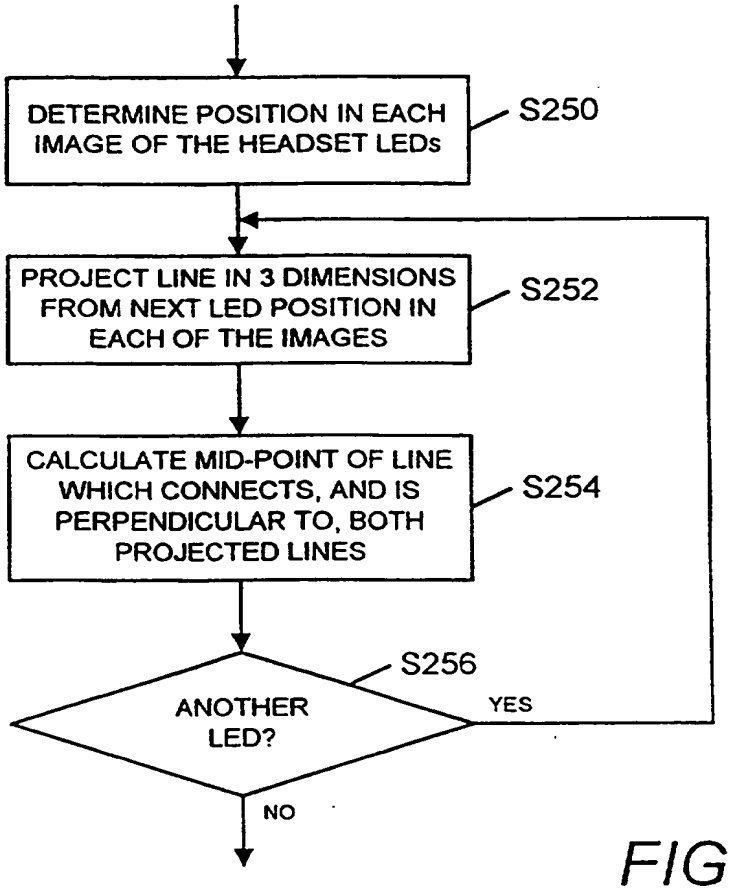

DETERMINE POSITION IN EACH IMAGE OF THE HEADSET LEDs
S250
PROJECT LINE IN 3 DIMENSIONS FROM NEXT LED POSITION IN EACH OF THE IMAGES
S252
CALCULATE MID-POINT OF LINE WHICH CONNECTS, AND IS PERPENDICULAR TO, BOTH PROJECTED LINES
S254
S256
ANOTHER LED?
YES
NO

FIG. 15

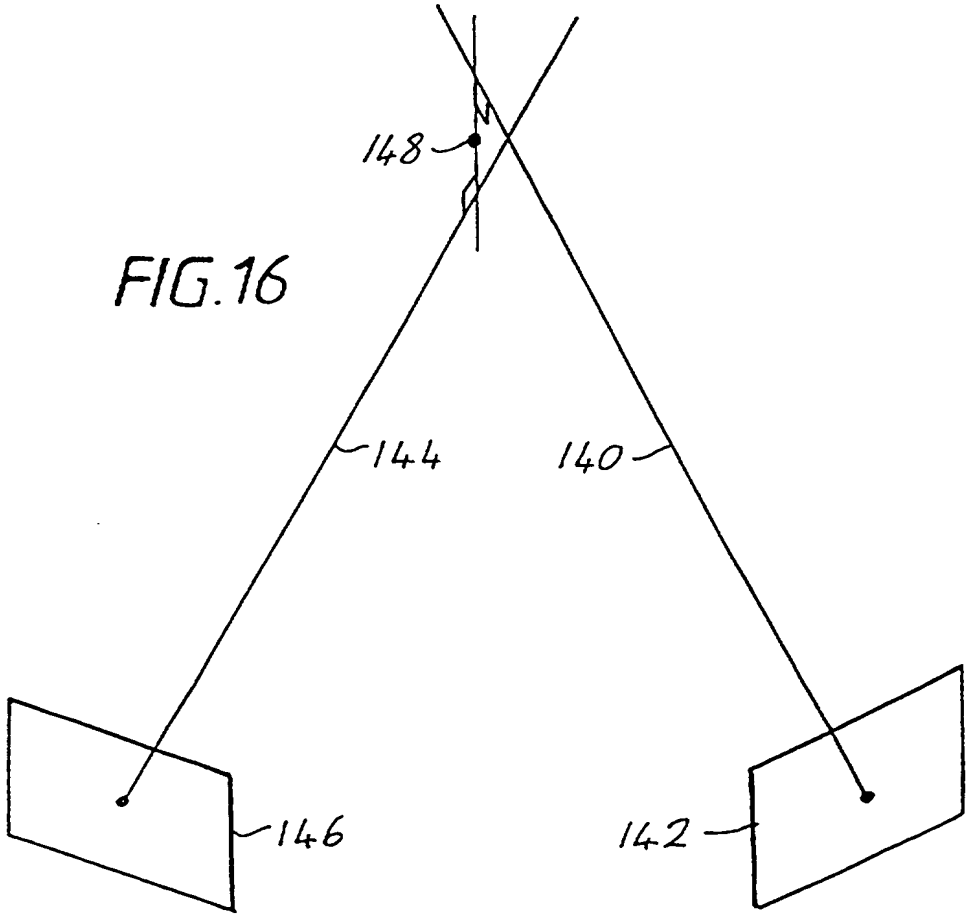

148
144
140
146
142

FIG. 16

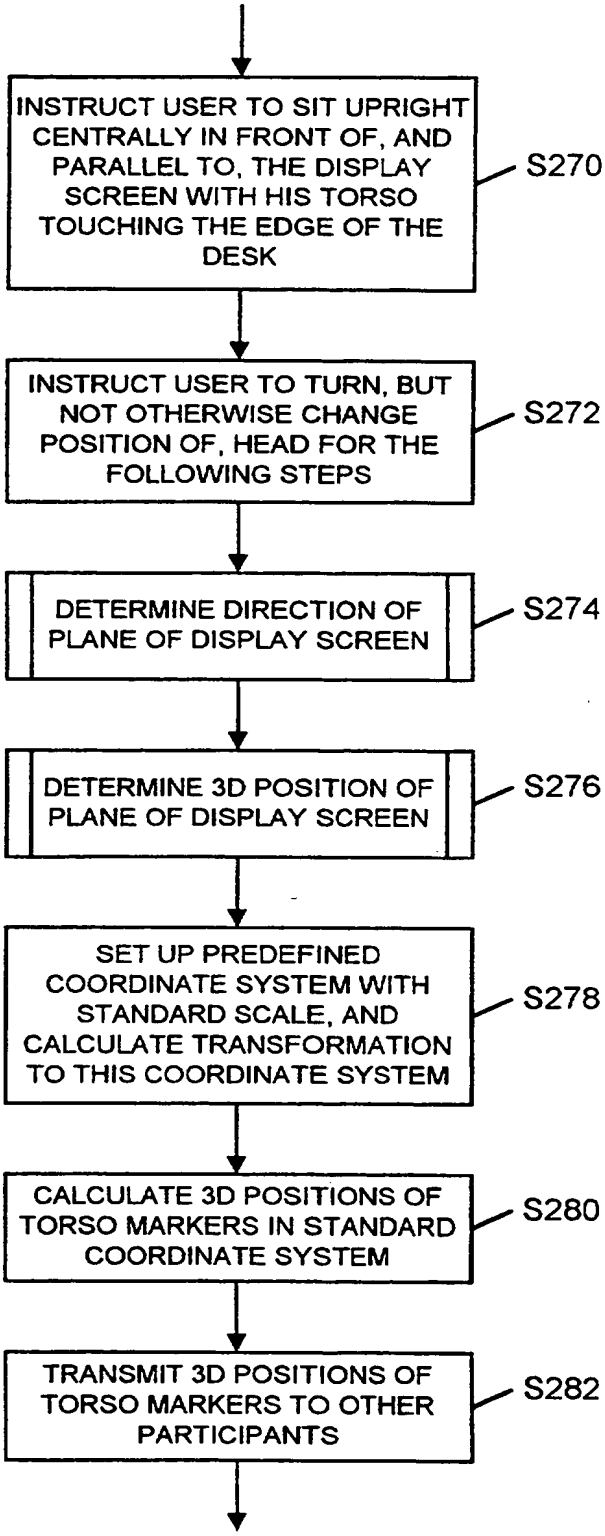

INSTRUCT USER TO SIT UPRIGHT CENTRALLY IN FRONT OF, AND PARALLEL TO, THE DISPLAY SCREEN WITH HIS TORSO TOUCHING THE EDGE OF THE DESK
S270
INSTRUCT USER TO TURN, BUT NOT OTHERWISE CHANGE POSITION OF, HEAD FOR THE FOLLOWING STEPS
S272
DETERMINE DIRECTION OF PLANE OF DISPLAY SCREEN
S274
DETERMINE 3D POSITION OF PLANE OF DISPLAY SCREEN
S276
SET UP PREDEFINED COORDINATE SYSTEM WITH STANDARD SCALE, AND CALCULATE TRANSFORMATION TO THIS COORDINATE SYSTEM
S278
CALCULATE 3D POSITIONS OF TORSO MARKERS IN STANDARD COORDINATE SYSTEM
S280
TRANSMIT 3D POSITIONS OF TORSO MARKERS TO OTHER PARTICIPANTS
S282

FIG. 17

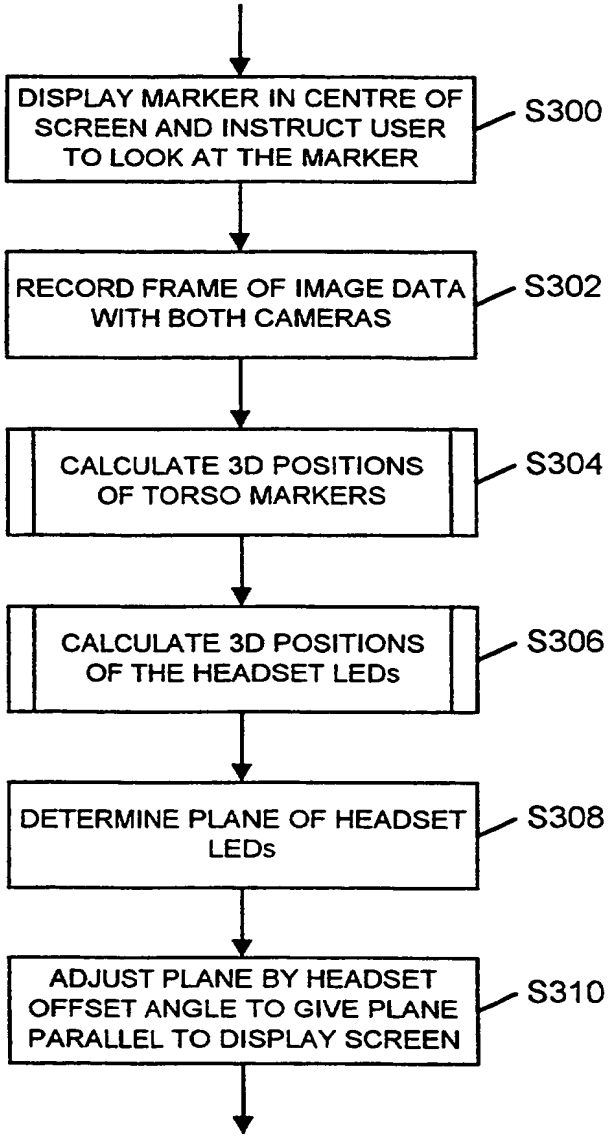

DISPLAY MARKER IN CENTRE OF SCREEN AND INSTRUCT USER TO LOOK AT THE MARKER
S300
RECORD FRAME OF IMAGE DATA WITH BOTH CAMERAS
S302
CALCULATE 3D POSITIONS OF TORSO MARKERS
S304
CALCULATE 3D POSITIONS OF THE HEADSET LEDs
S306
DETERMINE PLANE OF HEADSET LEDs
S308
ADJUST PLANE BY HEADSET OFFSET ANGLE TO GIVE PLANE PARALLEL TO DISPLAY SCREEN
S310

FIG. 18

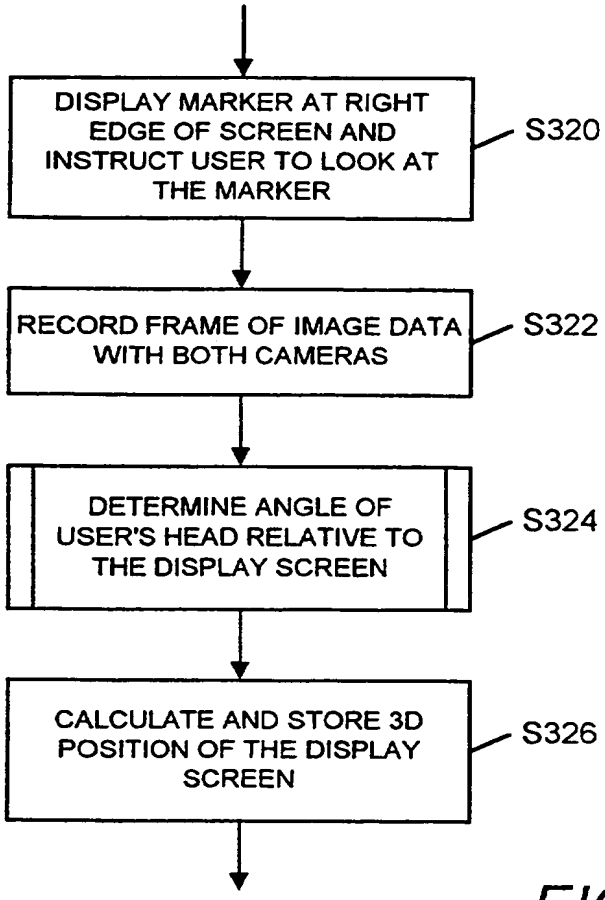

DISPLAY MARKER AT RIGHT EDGE OF SCREEN AND INSTRUCT USER TO LOOK AT THE MARKER
S320
RECORD FRAME OF IMAGE DATA WITH BOTH CAMERAS
S322
DETERMINE ANGLE OF USER'S HEAD RELATIVE TO THE DISPLAY SCREEN
S324
CALCULATE AND STORE 3D POSITION OF THE DISPLAY SCREEN
S326

FIG. 19

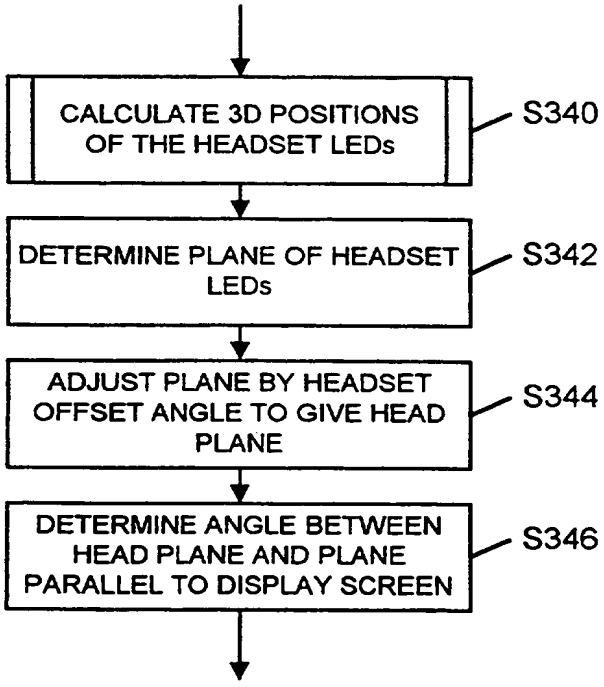
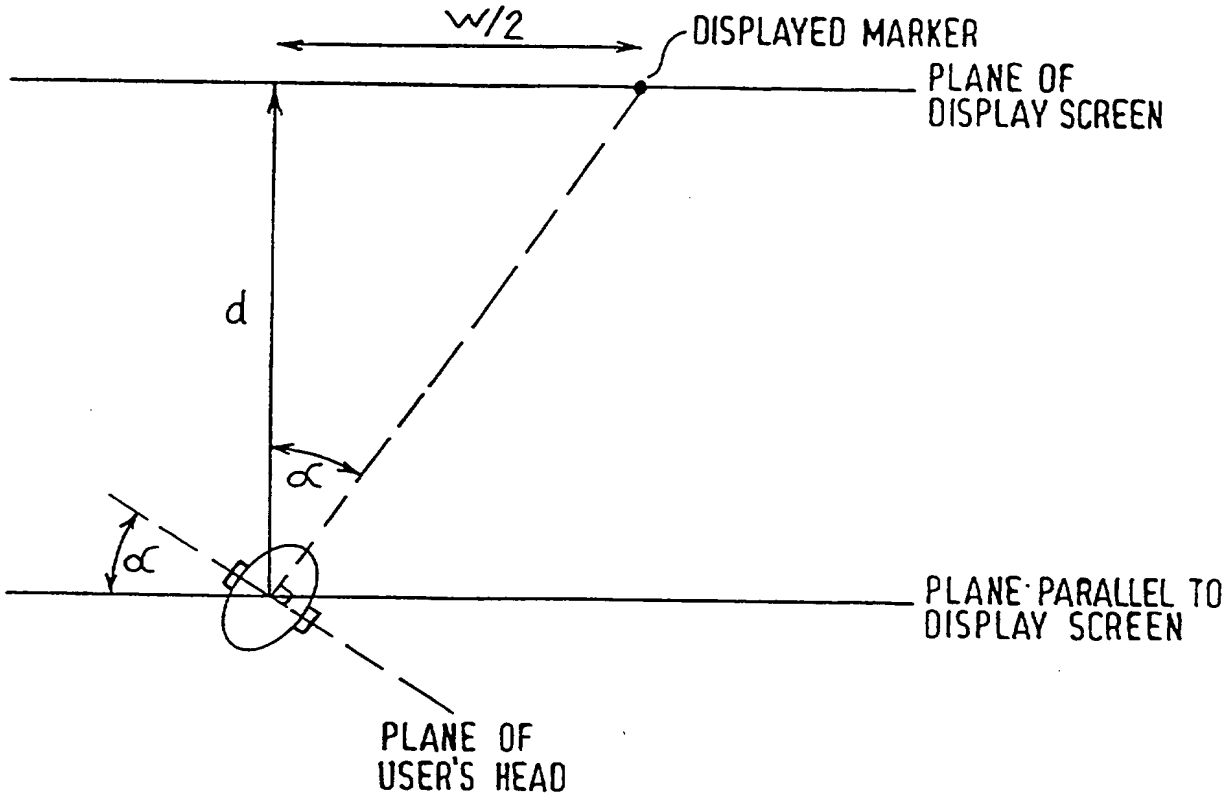

CALCULATE 3D POSITIONS OF THE HEADSET LEDs
S340
DETERMINE PLANE OF HEADSET LEDs
S342
ADJUST PLANE BY HEADSET OFFSET ANGLE TO GIVE HEAD PLANE
S344
DETERMINE ANGLE BETWEEN HEAD PLANE AND PLANE PARALLEL TO DISPLAY SCREEN
S346

FIG. 20

FIG. 21


W/2
DISPLAYED MARKER
PLANE OF
DISPLAY SCREEN
d
α
α
PLANE PARALLEL TO
DISPLAY SCREEN
PLANE OF
USER'S HEAD

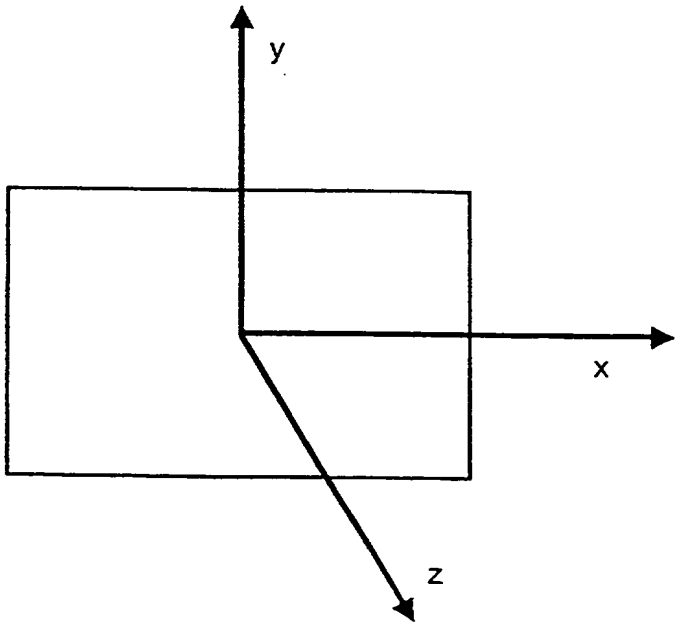

y
x
z

FIG. 22

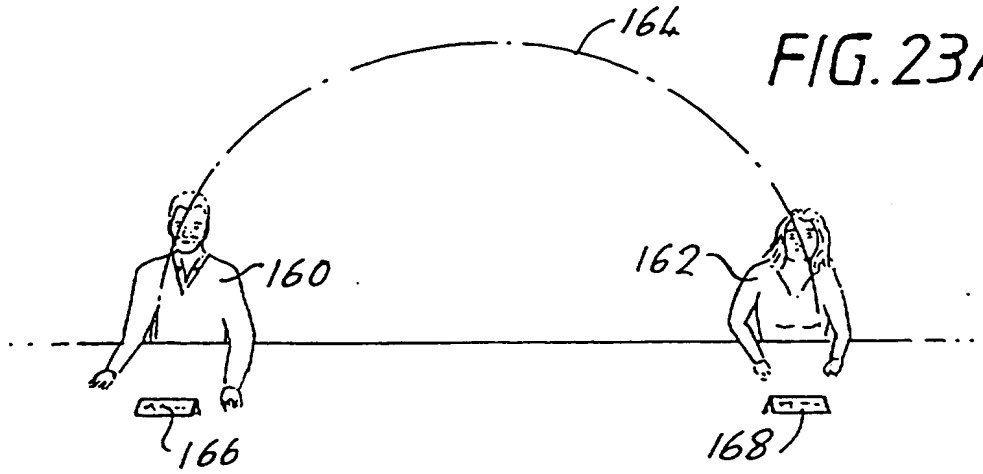


FIG. 23A

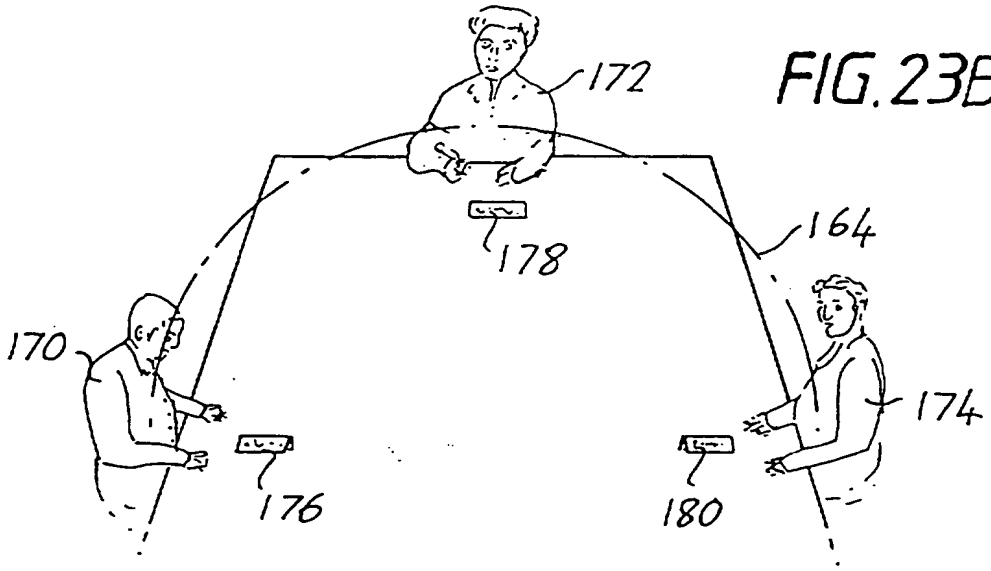


FIG. 23B

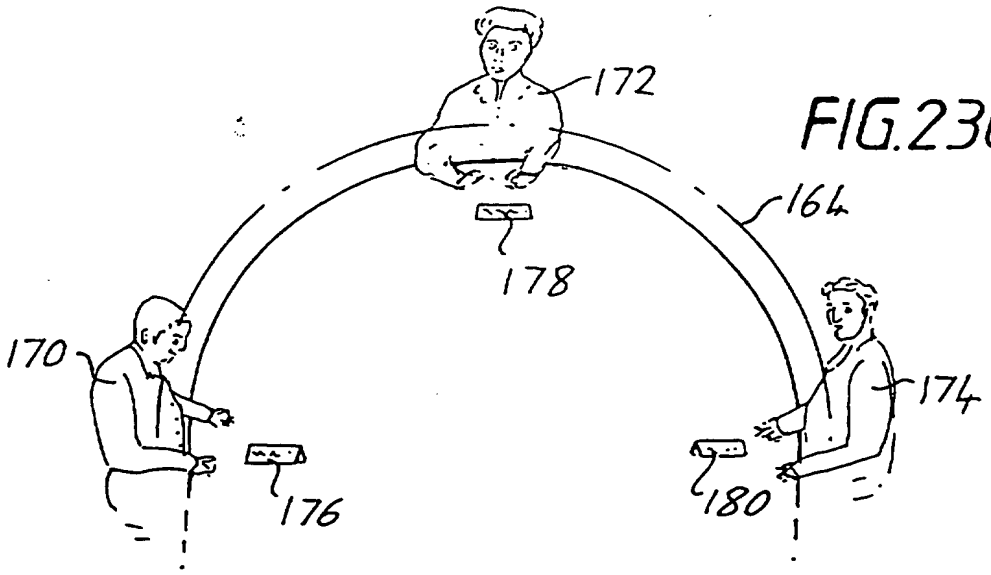


FIG. 23C

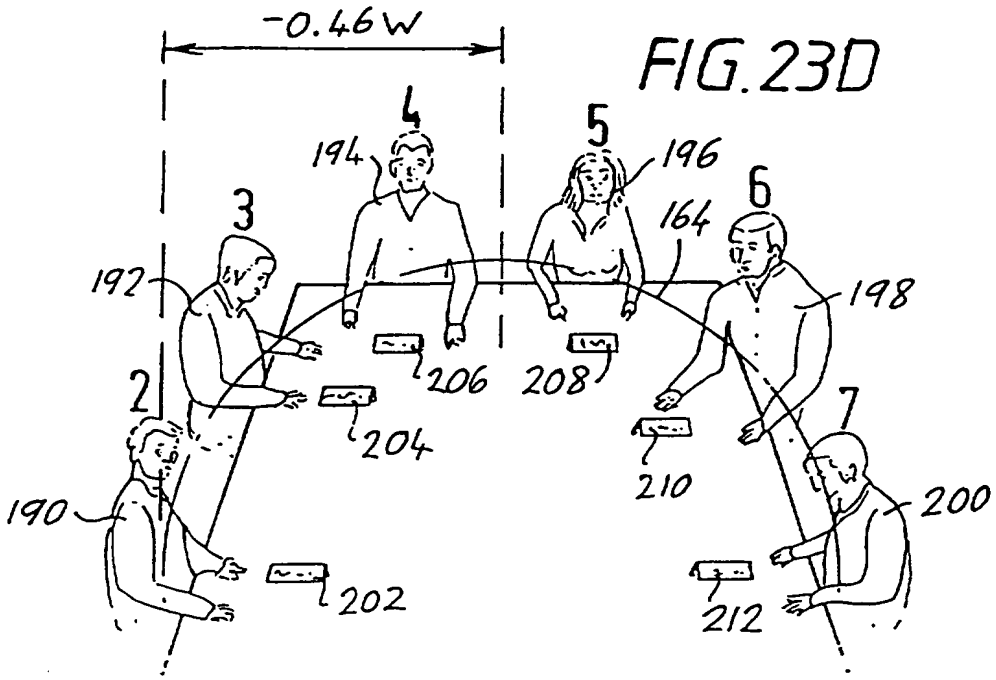


FIG. 23D

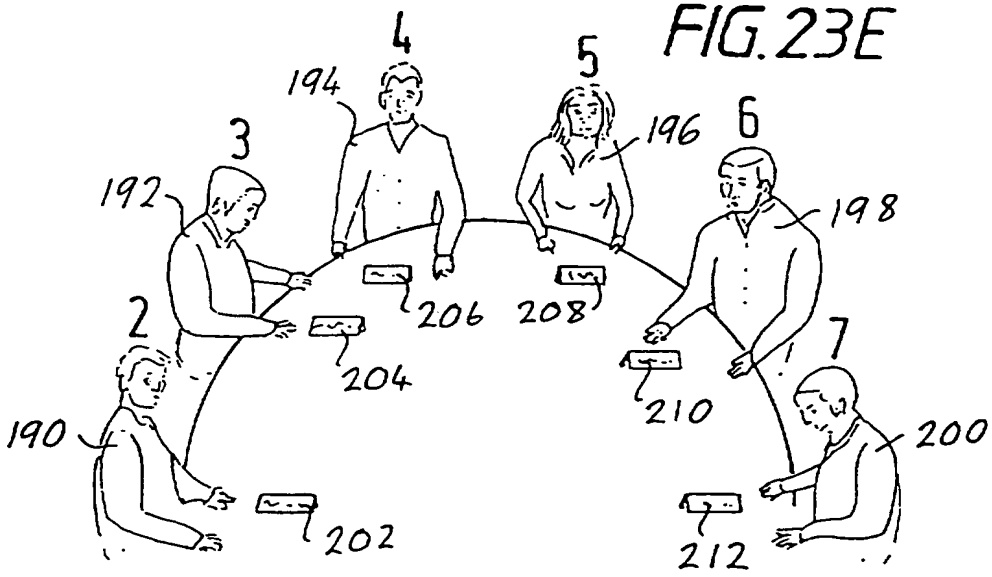


FIG. 23E

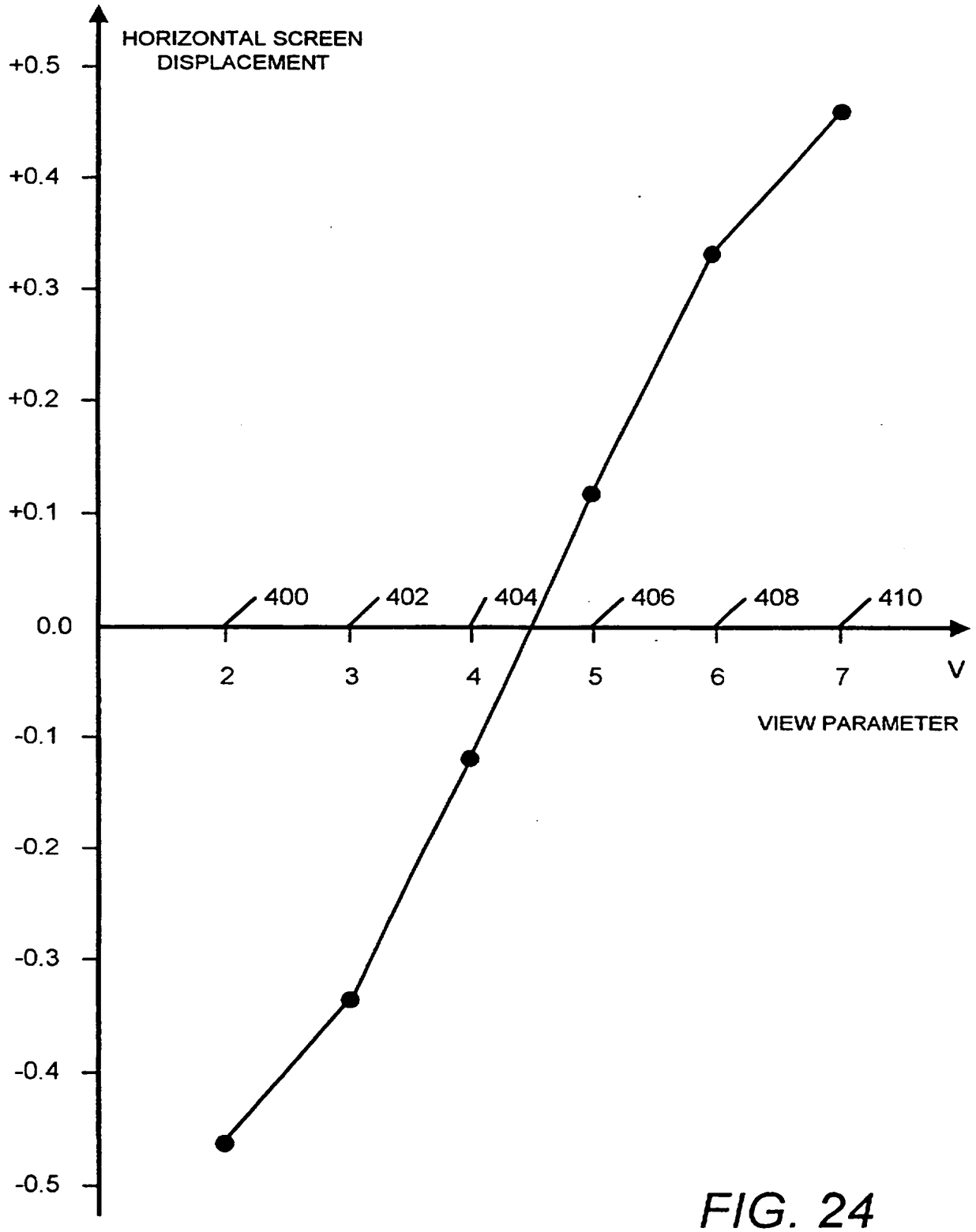

HORIZONTAL SCREEN DISPLACEMENT
+0.5
+0.4
+0.3
+0.2
+0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
400
402
404
406
408
410
2
3
4
5
6
7
V
VIEW PARAMETER

FIG. 24

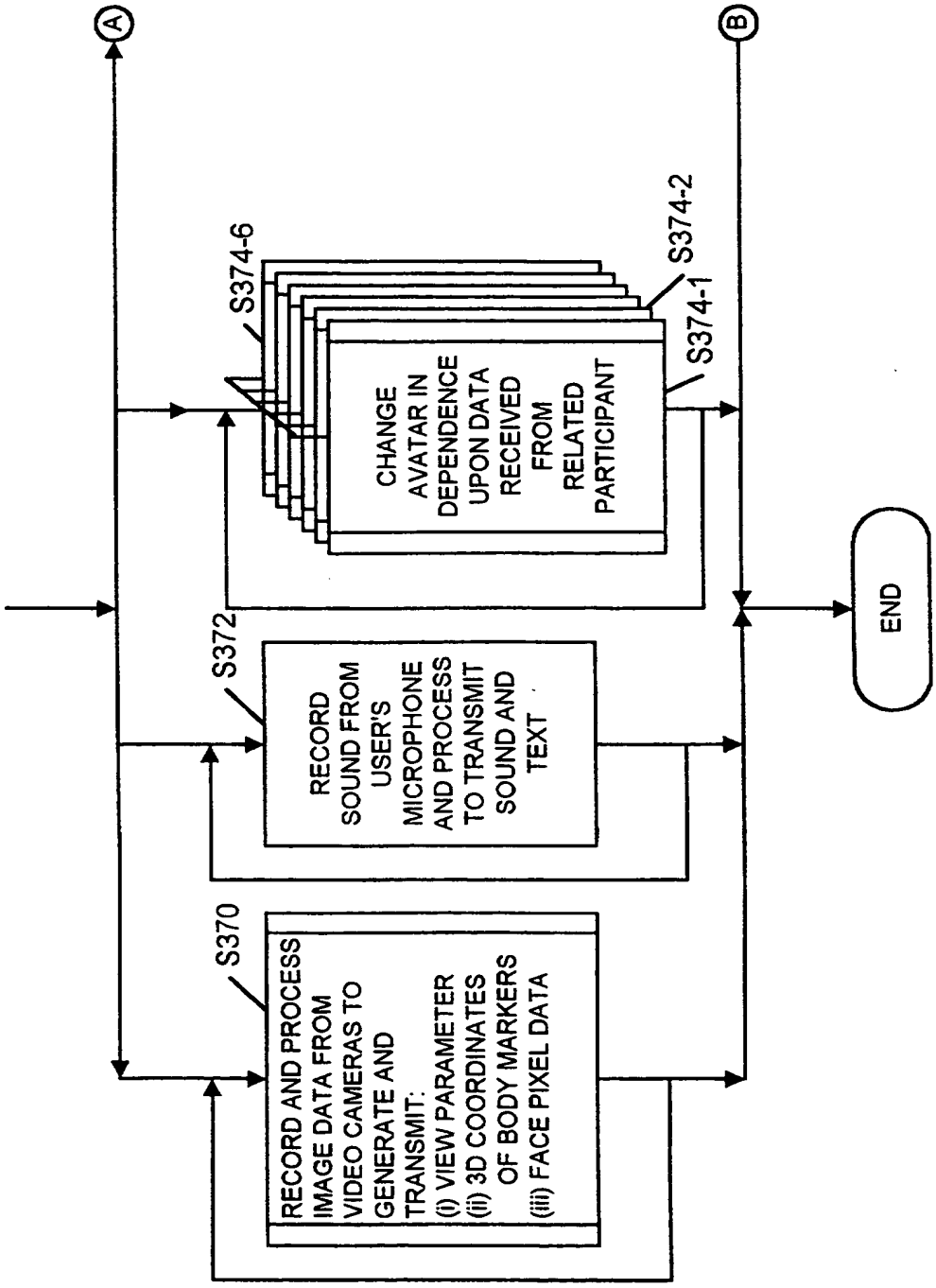

A
B
S374-6
S374-2
S374-1
CHANGE AVATAR IN DEPENDENCE UPON DATA RECEIVED FROM RELATED PARTICIPANT
S372
RECORD SOUND FROM USER'S MICROPHONE AND PROCESS TO TRANSMIT SOUND AND TEXT
S370
RECORD AND PROCESS IMAGE DATA FROM VIDEO CAMERAS TO GENERATE AND TRANSMIT:
(i) VIEW PARAMETER
(ii) 3D COORDINATES OF BODY MARKERS
(iii) FACE PIXEL DATA
END

FIG. 25

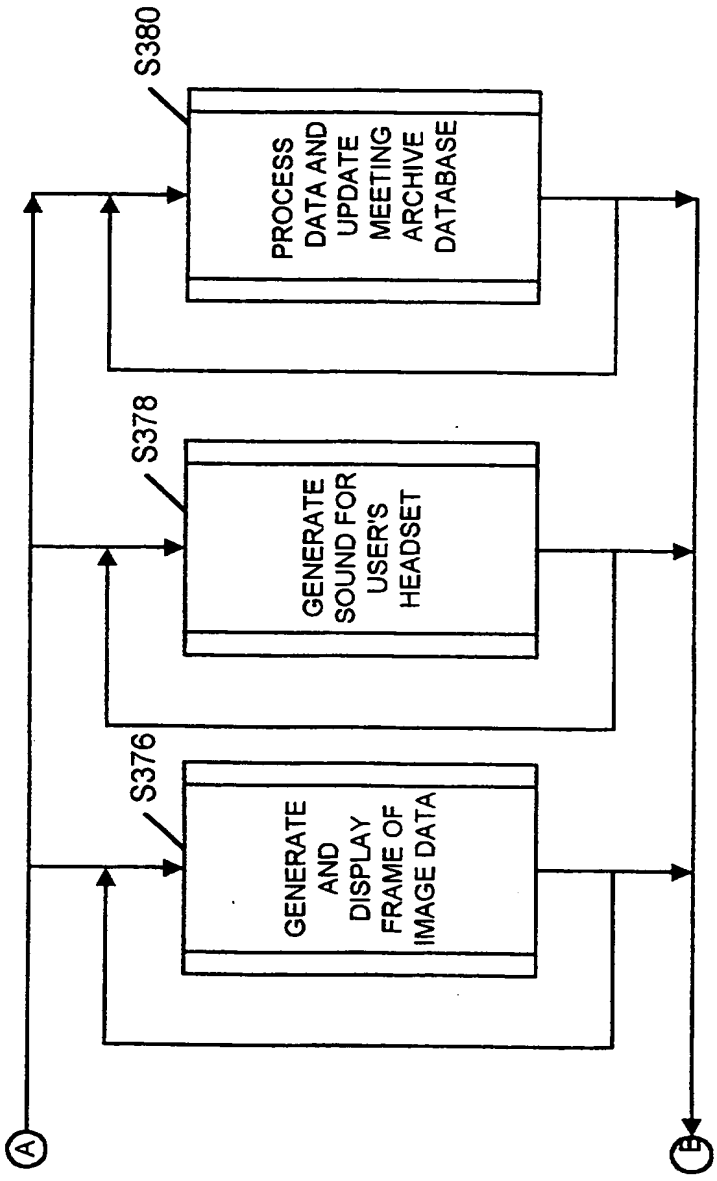

S380
PROCESS DATA AND UPDATE MEETING ARCHIVE DATABASE
S378
GENERATE SOUND FOR USER'S HEADSET
S376
GENERATE AND DISPLAY FRAME OF IMAGE DATA
A
B

FIG. 25 (cont)

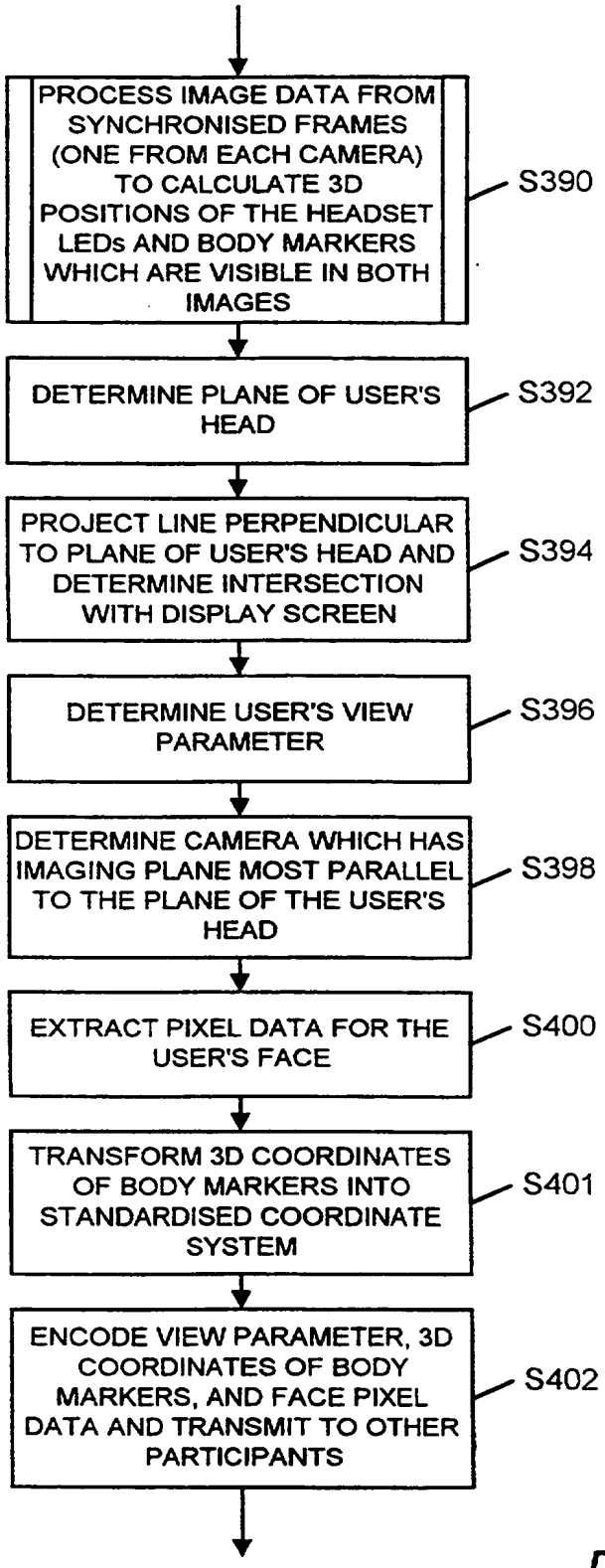

PROCESS IMAGE DATA FROM SYNCHRONISED FRAMES (ONE FROM EACH CAMERA) TO CALCULATE 3D POSITIONS OF THE HEADSET LEDs AND BODY MARKERS WHICH ARE VISIBLE IN BOTH IMAGES
S390
DETERMINE PLANE OF USER'S HEAD
S392
PROJECT LINE PERPENDICULAR TO PLANE OF USER'S HEAD AND DETERMINE INTERSECTION WITH DISPLAY SCREEN
S394
DETERMINE USER'S VIEW PARAMETER
S396
DETERMINE CAMERA WHICH HAS IMAGING PLANE MOST PARALLEL TO THE PLANE OF THE USER'S HEAD
S398
EXTRACT PIXEL DATA FOR THE USER'S FACE
S400
TRANSFORM 3D COORDINATES OF BODY MARKERS INTO STANDARDISED COORDINATE SYSTEM
S401
ENCODE VIEW PARAMETER, 3D COORDINATES OF BODY MARKERS, AND FACE PIXEL DATA AND TRANSMIT TO OTHER PARTICIPANTS
S402

FIG. 26

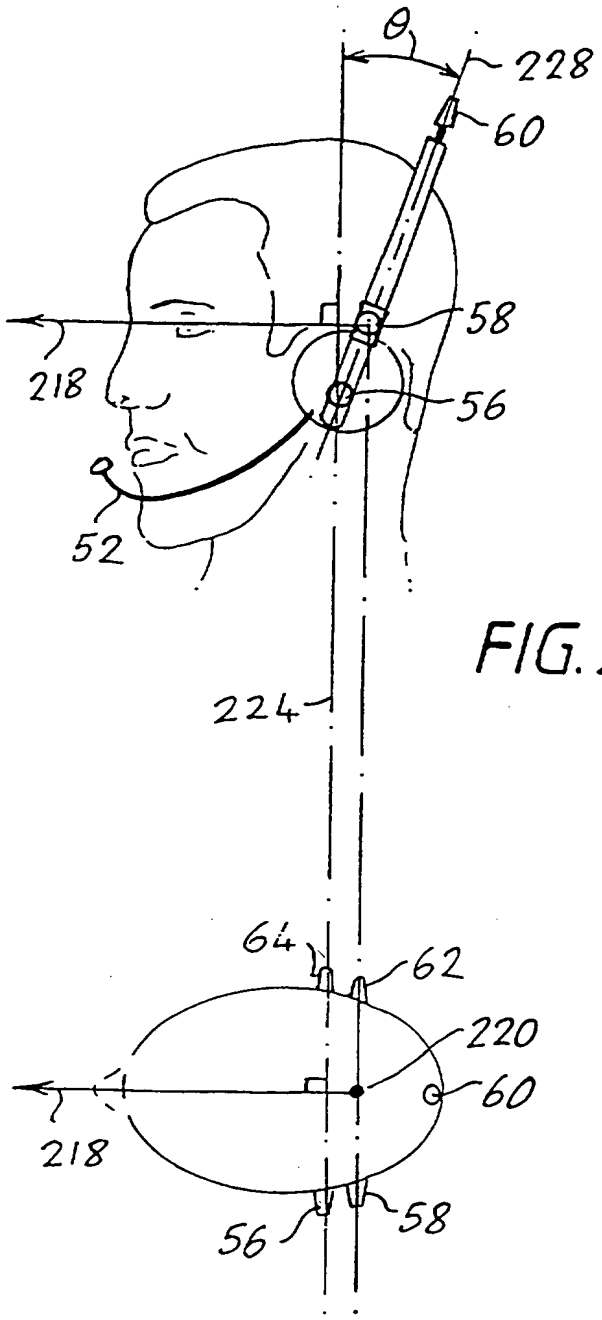


FIG. 27A

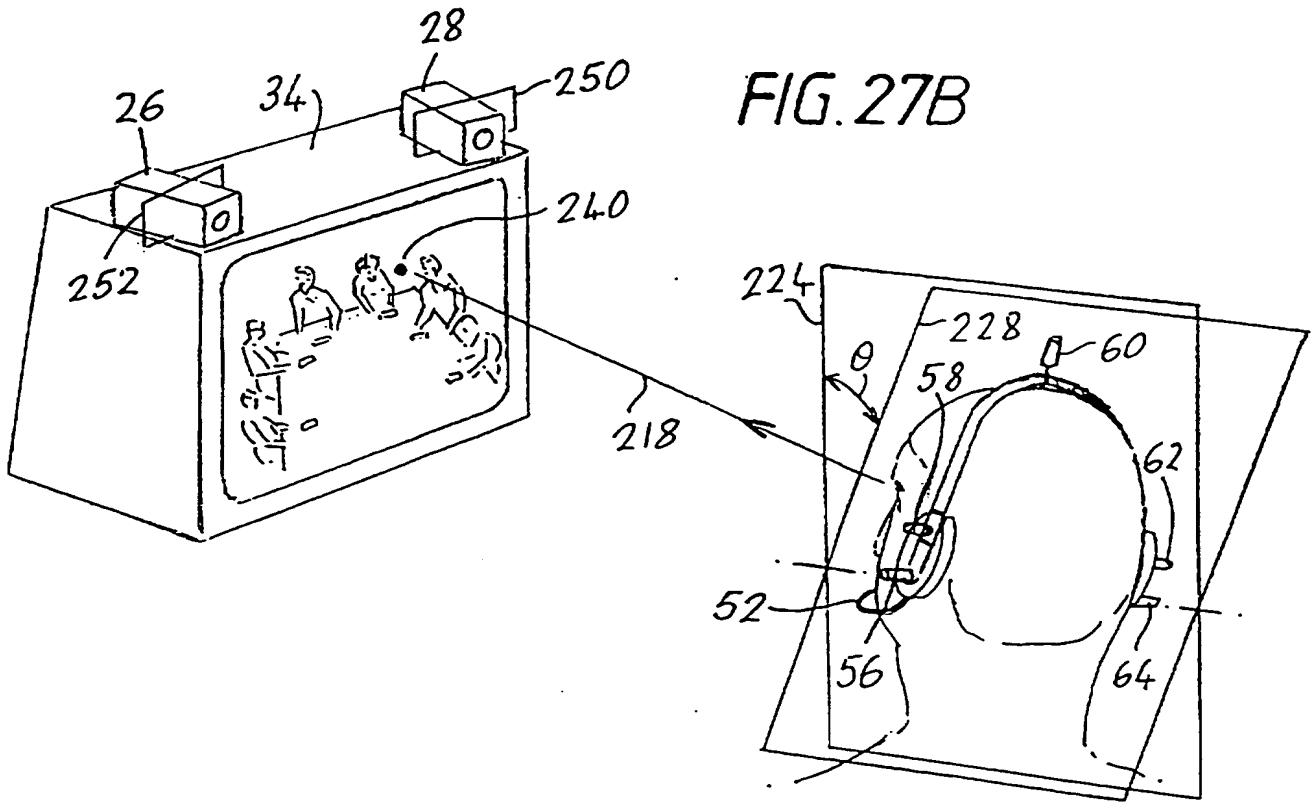
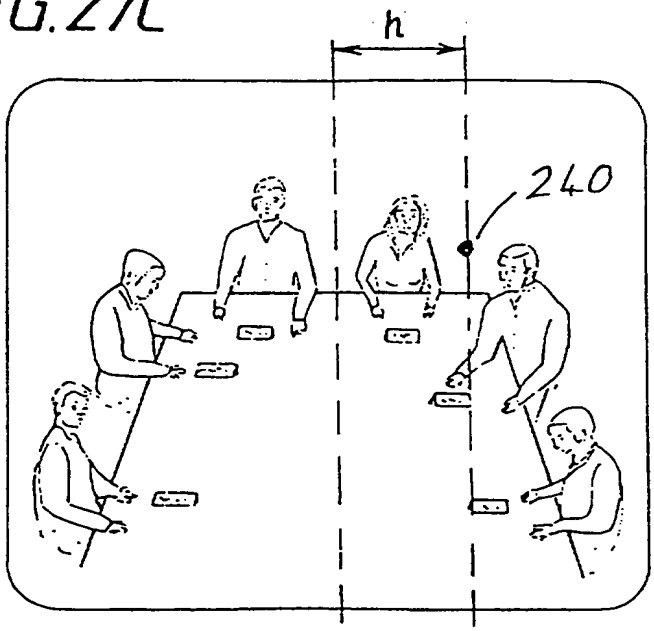


FIG. 27B

FIG. 27C



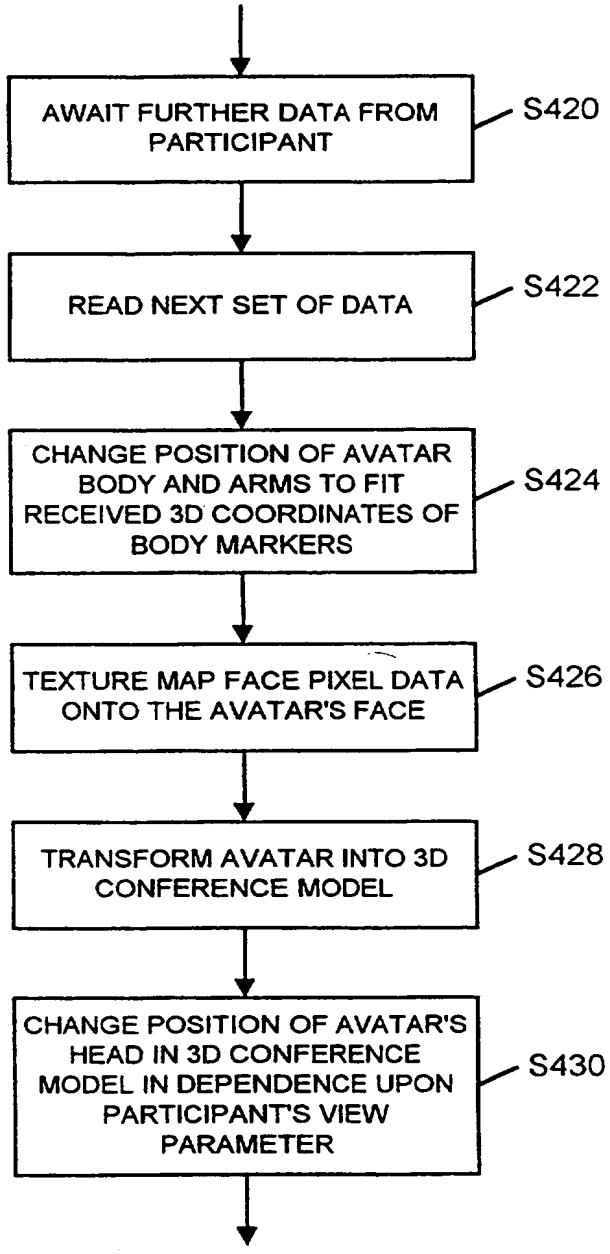

AWAIT FURTHER DATA FROM PARTICIPANT
S420
READ NEXT SET OF DATA
S422
CHANGE POSITION OF AVATAR BODY AND ARMS TO FIT RECEIVED 3D COORDINATES OF BODY MARKERS
S424
TEXTURE MAP FACE PIXEL DATA ONTO THE AVATAR'S FACE
S426
TRANSFORM AVATAR INTO 3D CONFERENCE MODEL
S428
CHANGE POSITION OF AVATAR'S HEAD IN 3D CONFERENCE MODEL IN DEPENDENCE UPON PARTICIPANT'S VIEW PARAMETER
S430

FIG. 28

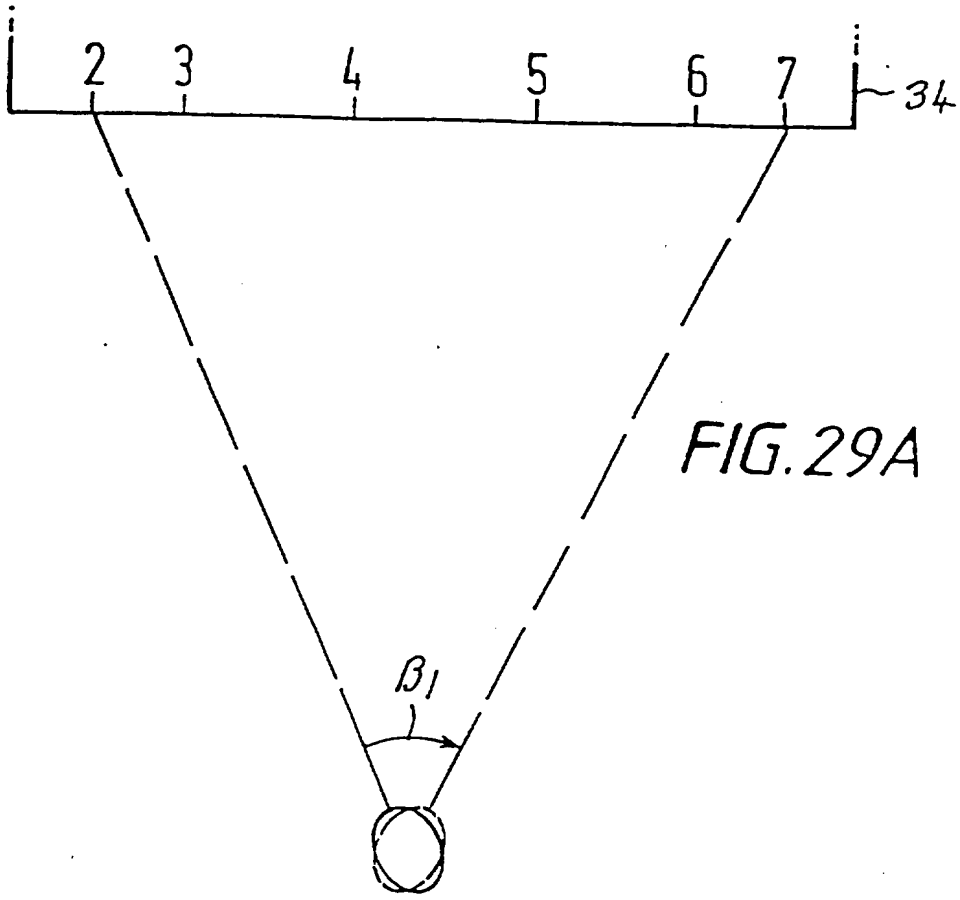


FIG. 29A

FIG. 29B

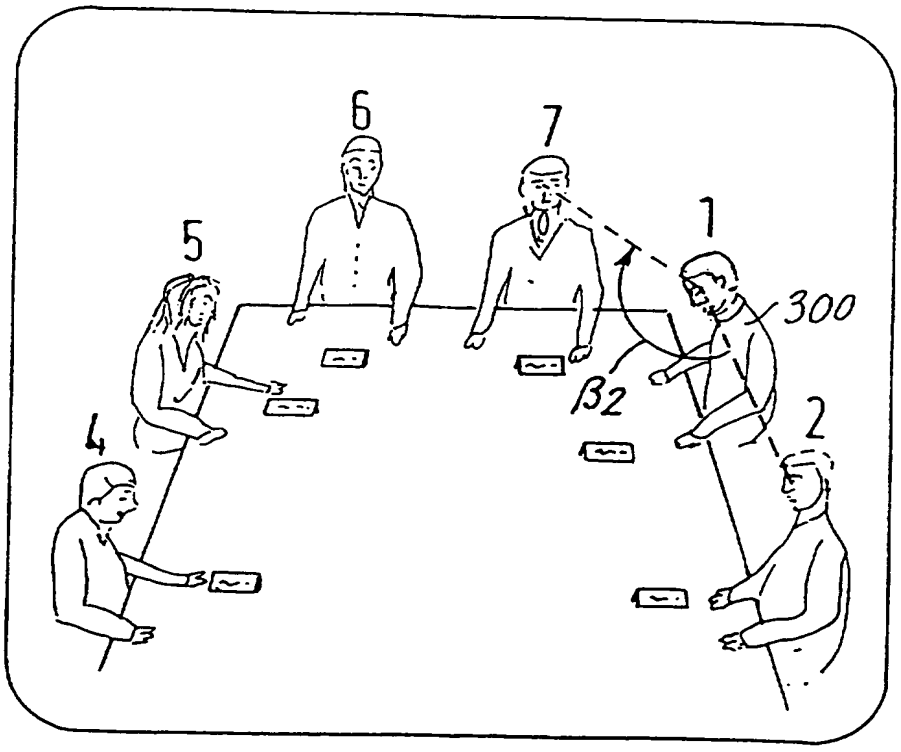
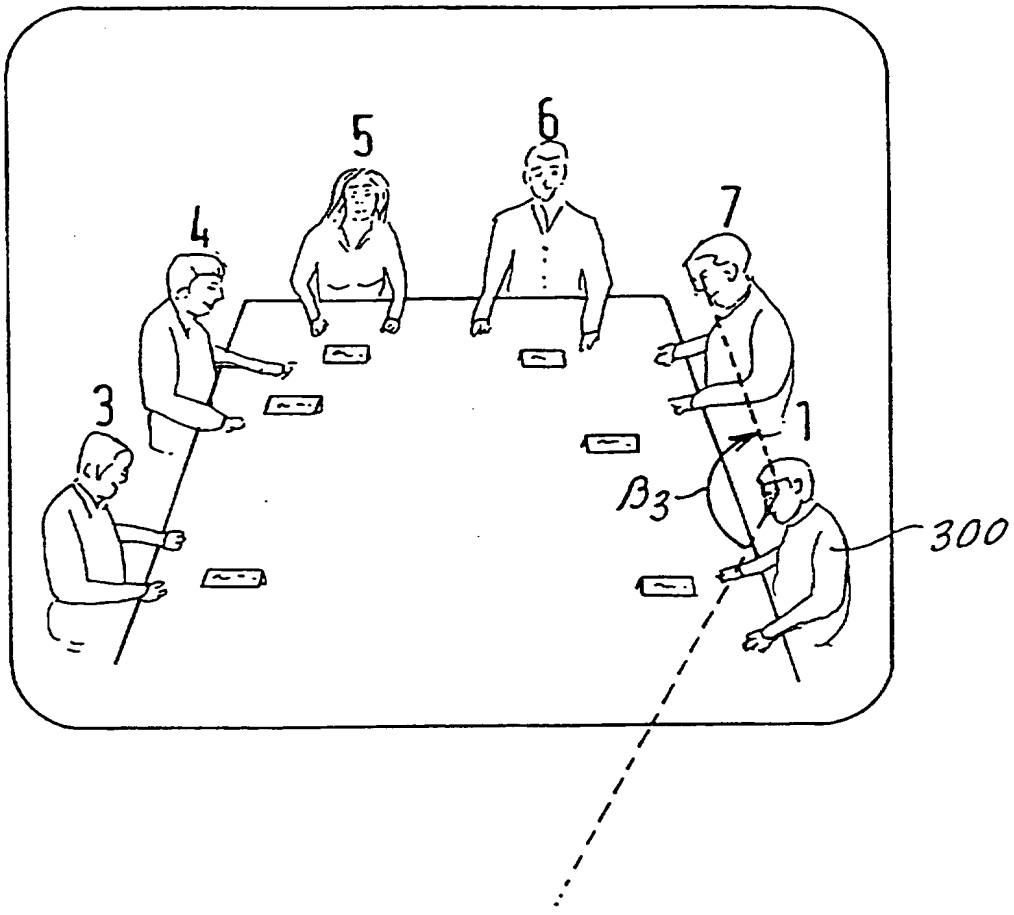


FIG. 29C



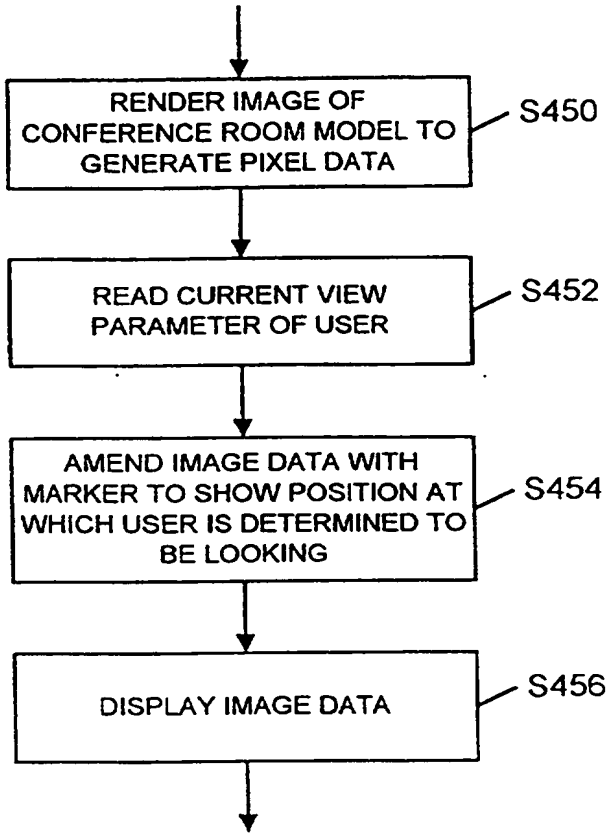

RENDER IMAGE OF CONFERENCE ROOM MODEL TO GENERATE PIXEL DATA
S450
READ CURRENT VIEW PARAMETER OF USER
S452
AMEND IMAGE DATA WITH MARKER TO SHOW POSITION AT WHICH USER IS DETERMINED TO BE LOOKING
S454
DISPLAY IMAGE DATA
S456

FIG. 30

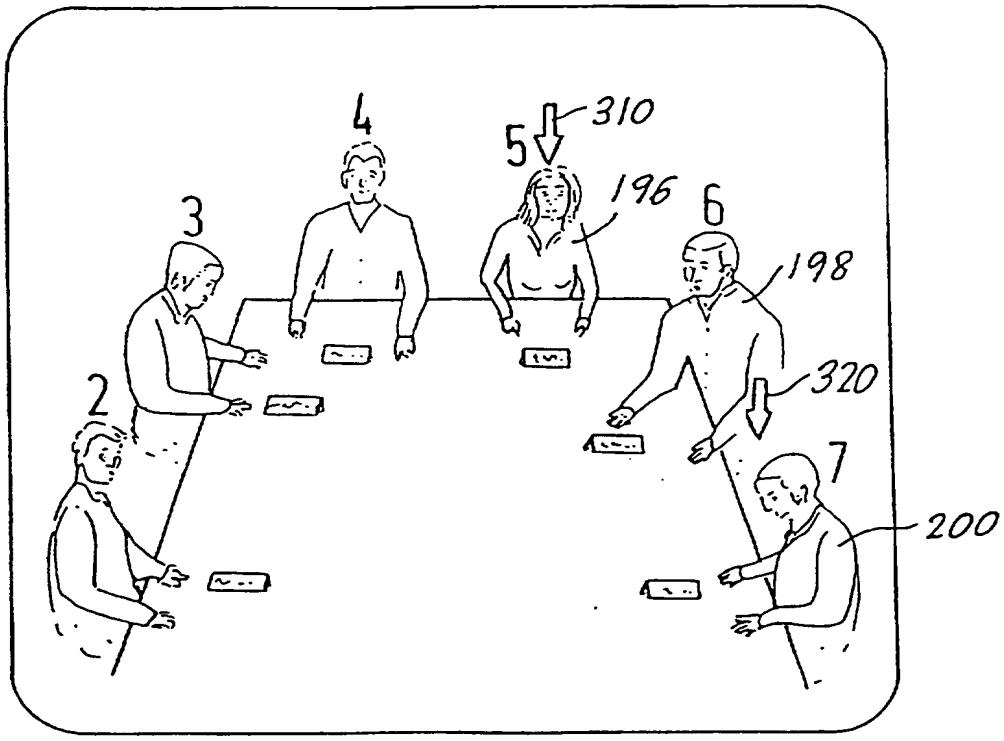

2
3
4
5
310
196
6
198
320
7
200

FIG. 31

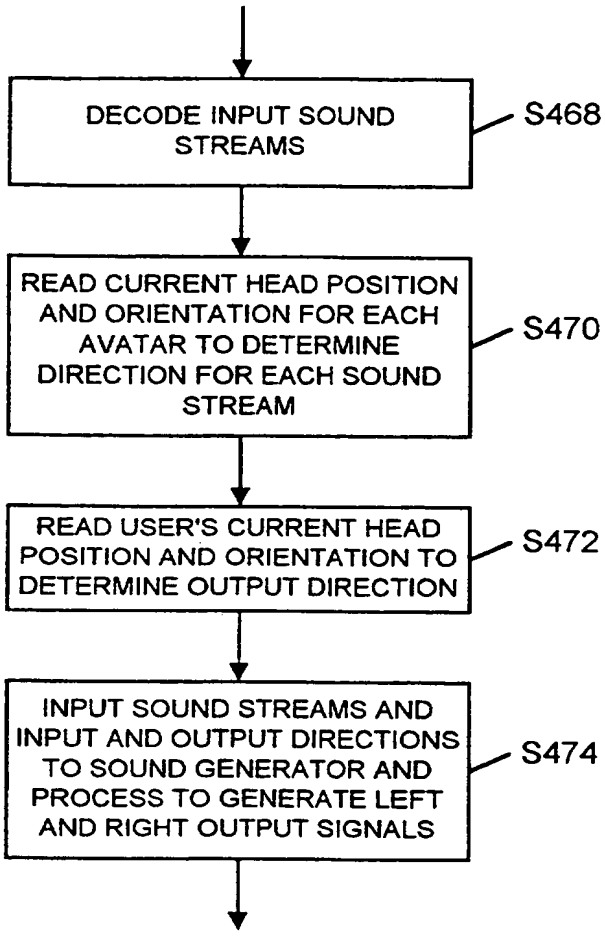

DECODE INPUT SOUND STREAMS
S468
READ CURRENT HEAD POSITION AND ORIENTATION FOR EACH AVATAR TO DETERMINE DIRECTION FOR EACH SOUND STREAM
S470
READ USER'S CURRENT HEAD POSITION AND ORIENTATION TO DETERMINE OUTPUT DIRECTION
S472
INPUT SOUND STREAMS AND INPUT AND OUTPUT DIRECTIONS TO SOUND GENERATOR AND PROCESS TO GENERATE LEFT AND RIGHT OUTPUT SIGNALS
S474

FIG. 32

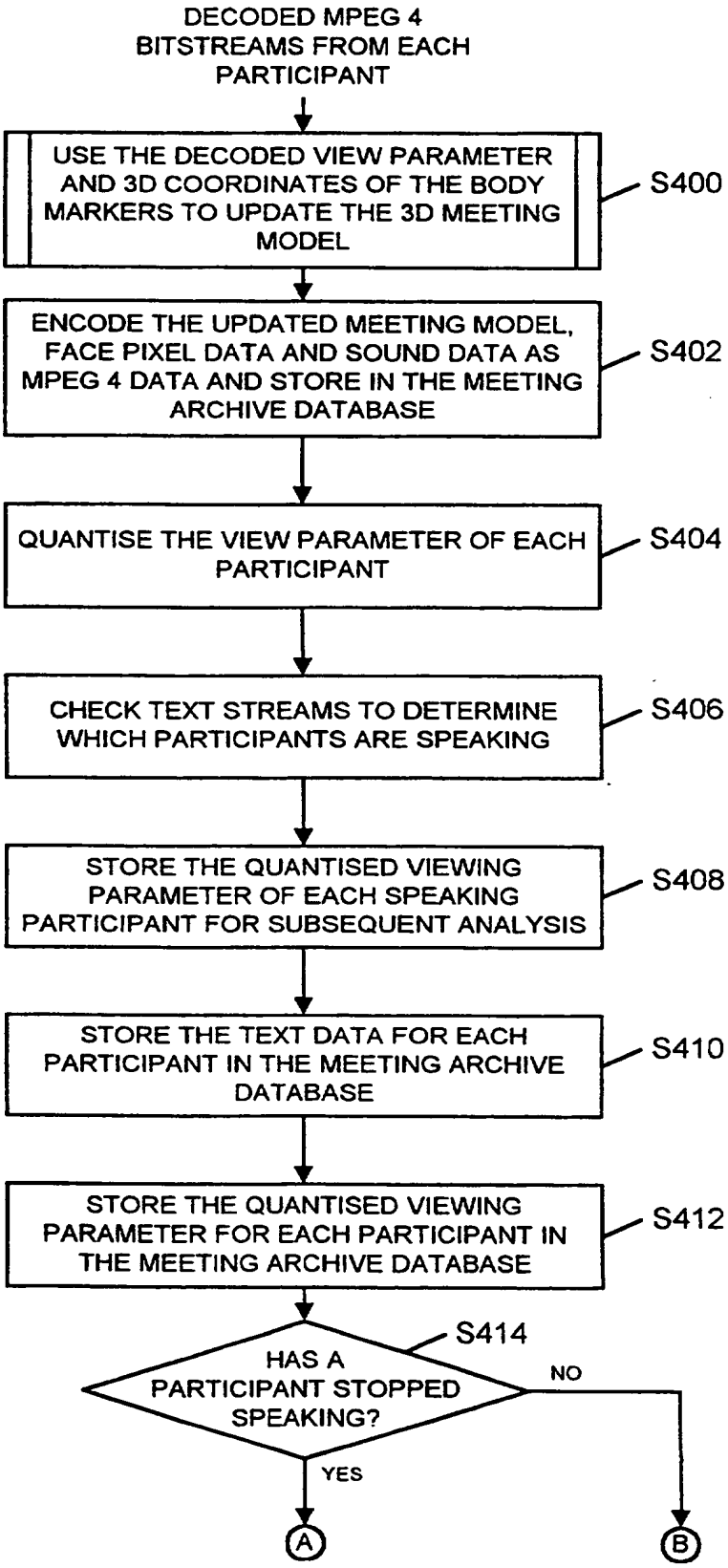

DECODED MPEG 4 BITSTREAMS FROM EACH PARTICIPANT
USE THE DECODED VIEW PARAMETER AND 3D COORDINATES OF THE BODY MARKERS TO UPDATE THE 3D MEETING MODEL
S400
ENCODE THE UPDATED MEETING MODEL, FACE PIXEL DATA AND SOUND DATA AS MPEG 4 DATA AND STORE IN THE MEETING ARCHIVE DATABASE
S402
QUANTISE THE VIEW PARAMETER OF EACH PARTICIPANT
S404
CHECK TEXT STREAMS TO DETERMINE WHICH PARTICIPANTS ARE SPEAKING
S406
STORE THE QUANTISED VIEWING PARAMETER OF EACH SPEAKING PARTICIPANT FOR SUBSEQUENT ANALYSIS
S408
STORE THE TEXT DATA FOR EACH PARTICIPANT IN THE MEETING ARCHIVE DATABASE
S410
STORE THE QUANTISED VIEWING PARAMETER FOR EACH PARTICIPANT IN THE MEETING ARCHIVE DATABASE
S412
HAS A PARTICIPANT STOPPED SPEAKING?
S414
NO
YES
A
B

FIG. 33

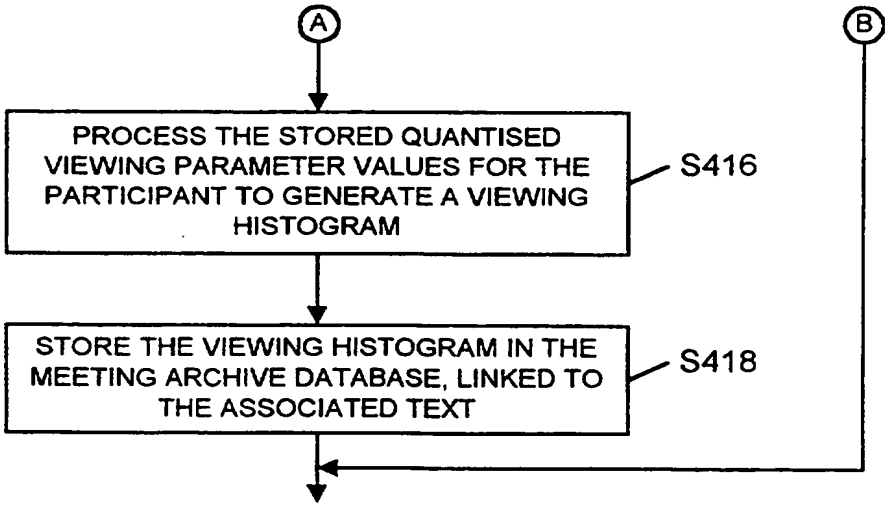

A
B
PROCESS THE STORED QUANTISED VIEWING PARAMETER VALUES FOR THE PARTICIPANT TO GENERATE A VIEWING HISTOGRAM
S416
STORE THE VIEWING HISTOGRAM IN THE MEETING ARCHIVE DATABASE, LINKED TO THE ASSOCIATED TEXT
S418

FIG. 33 (cont)

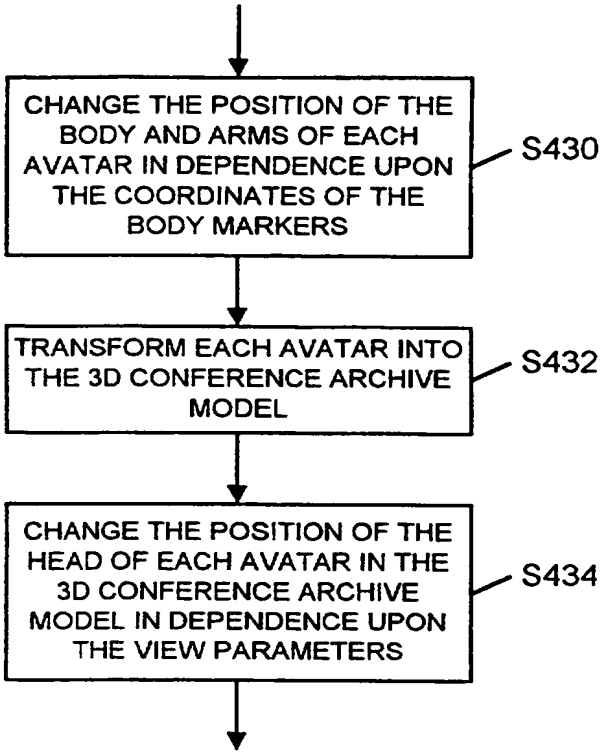

CHANGE THE POSITION OF THE BODY AND ARMS OF EACH AVATAR IN DEPENDENCE UPON THE COORDINATES OF THE BODY MARKERS
S430
TRANSFORM EACH AVATAR INTO THE 3D CONFERENCE ARCHIVE MODEL
S432
CHANGE THE POSITION OF THE HEAD OF EACH AVATAR IN THE 3D CONFERENCE ARCHIVE MODEL IN DEPENDENCE UPON THE VIEW PARAMETERS
S434

FIG. 34

FIG. 35

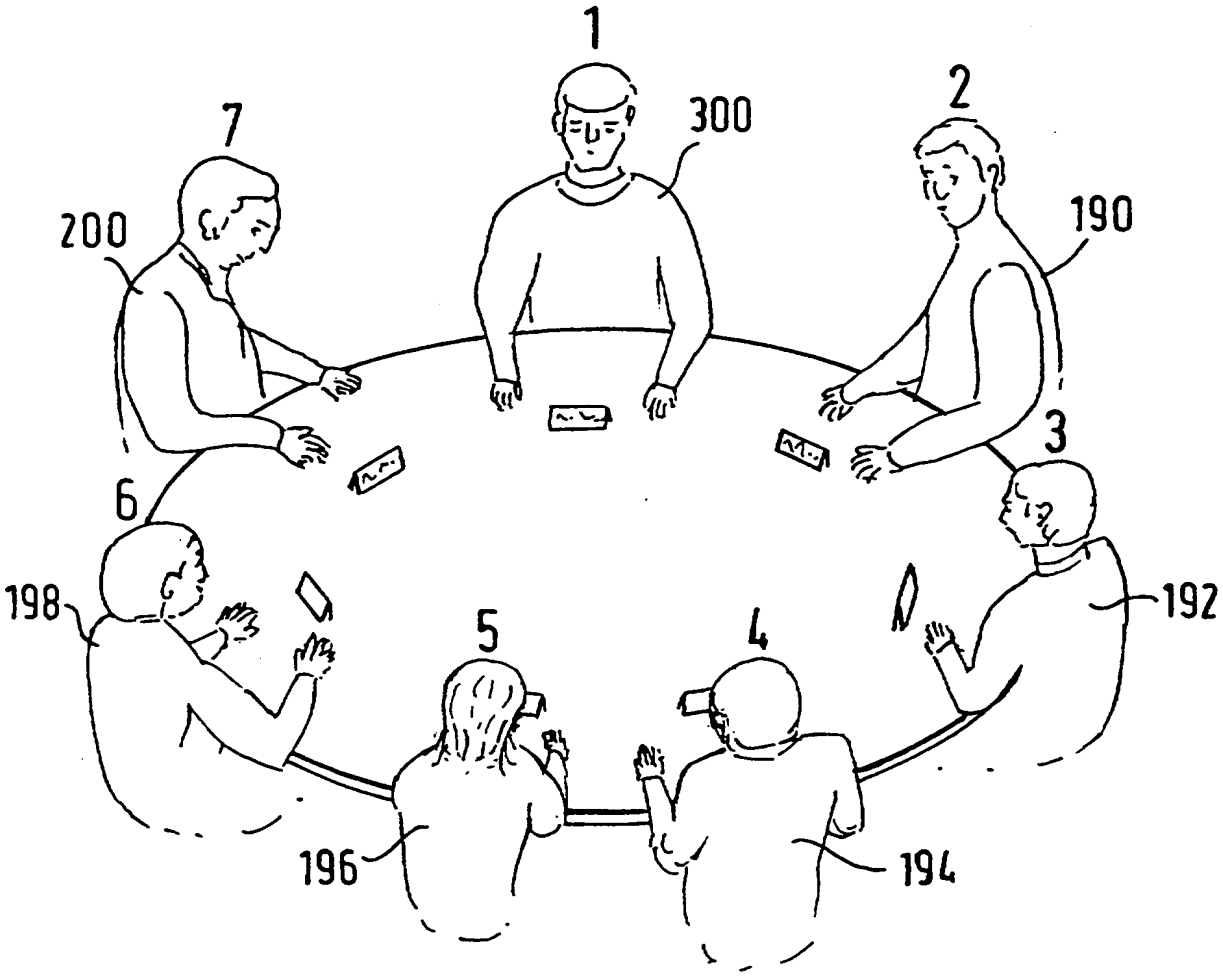


FIG. 36A

NUMBER	PARTICIPANT
1	MR. A
2	MR. B
3	MR. C
4	MR. D
5	MISS. E
6	MR. F
7	MR. G

480

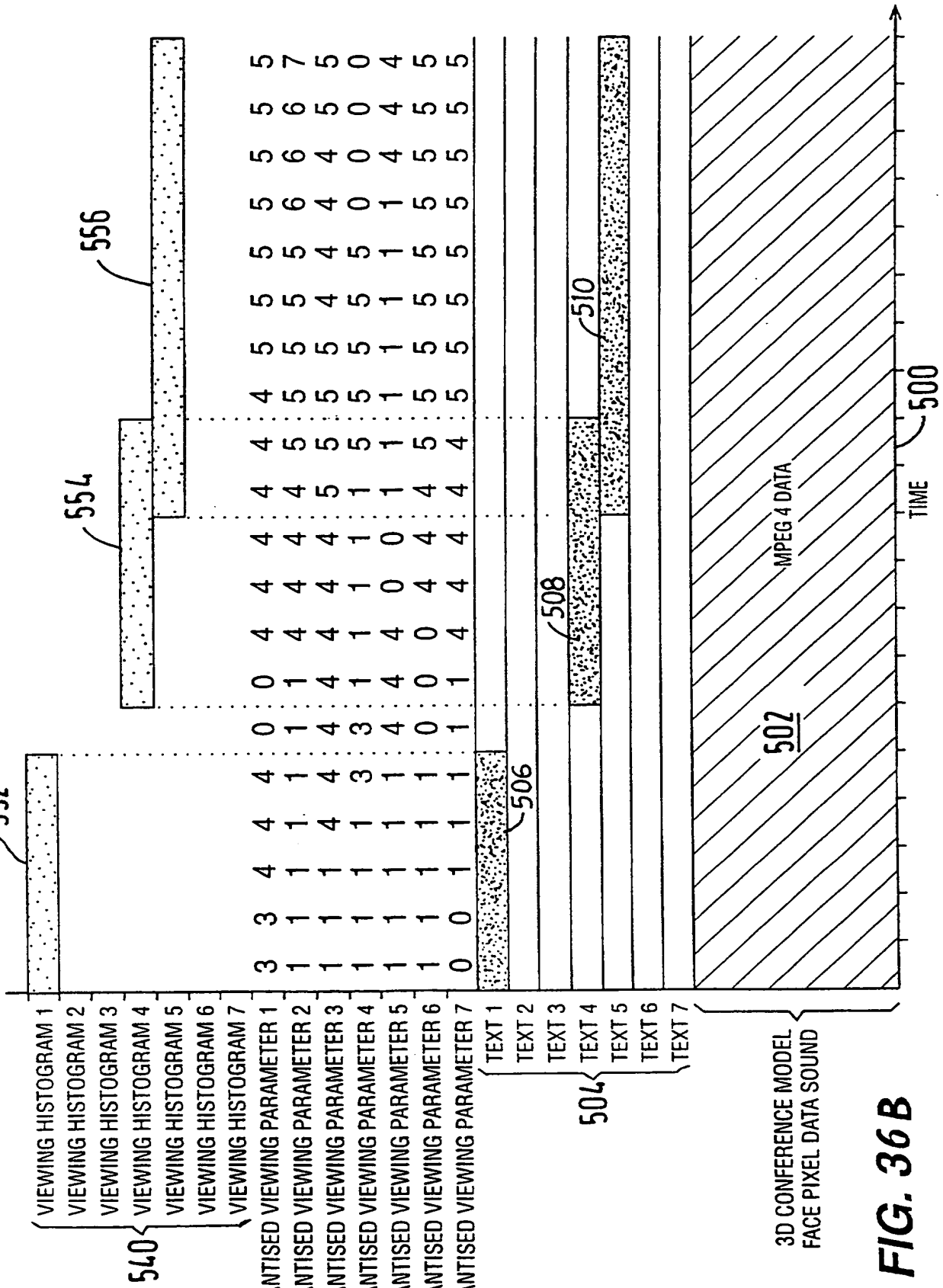

552
554
556
540
VIEWING HISTOGRAM 1
VIEWING HISTOGRAM 2
VIEWING HISTOGRAM 3
VIEWING HISTOGRAM 4
VIEWING HISTOGRAM 5
VIEWING HISTOGRAM 6
VIEWING HISTOGRAM 7
512
QUANTISED VIEWING PARAMETER 1
QUANTISED VIEWING PARAMETER 2
QUANTISED VIEWING PARAMETER 3
QUANTISED VIEWING PARAMETER 4
QUANTISED VIEWING PARAMETER 5
QUANTISED VIEWING PARAMETER 6
QUANTISED VIEWING PARAMETER 7
506
508
510
504
TEXT 1
TEXT 2
TEXT 3
TEXT 4
TEXT 5
TEXT 6
TEXT 7
502
MPEG 4 DATA
3D CONFERENCE MODEL
FACE PIXEL DATA SOUND
TIME
500

FIG. 36B

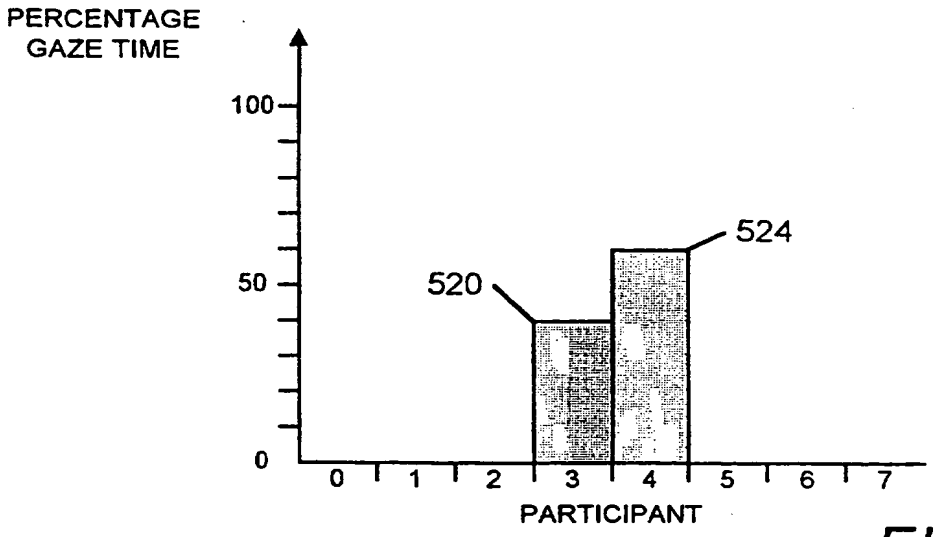


FIG. 37A

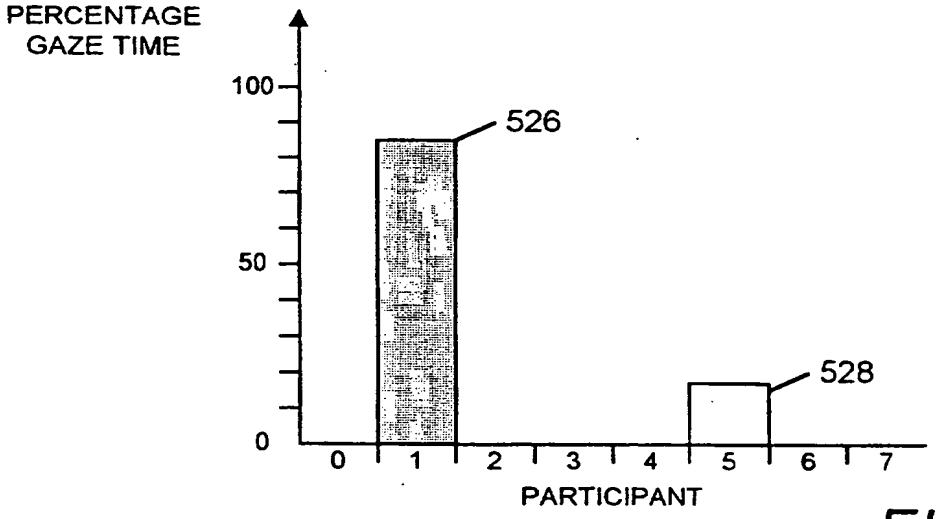


FIG. 37B

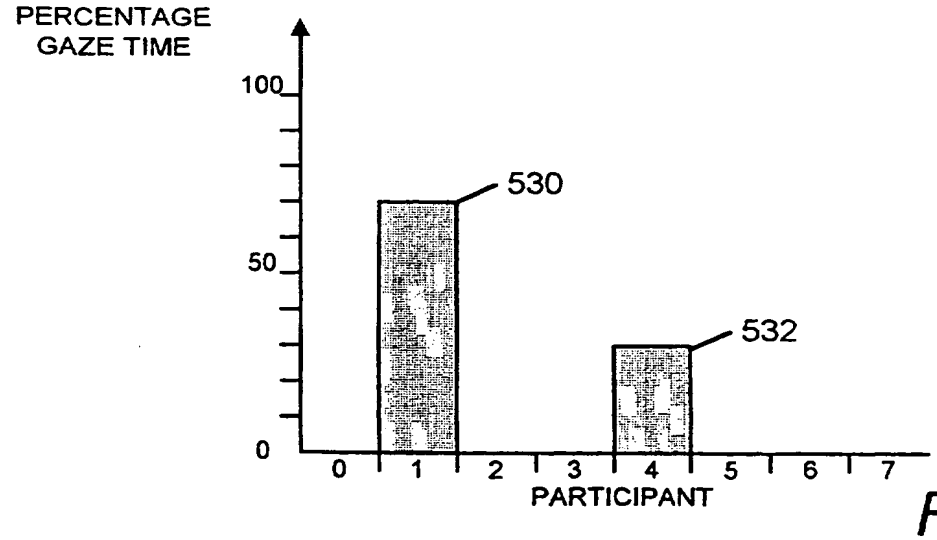


FIG. 37C

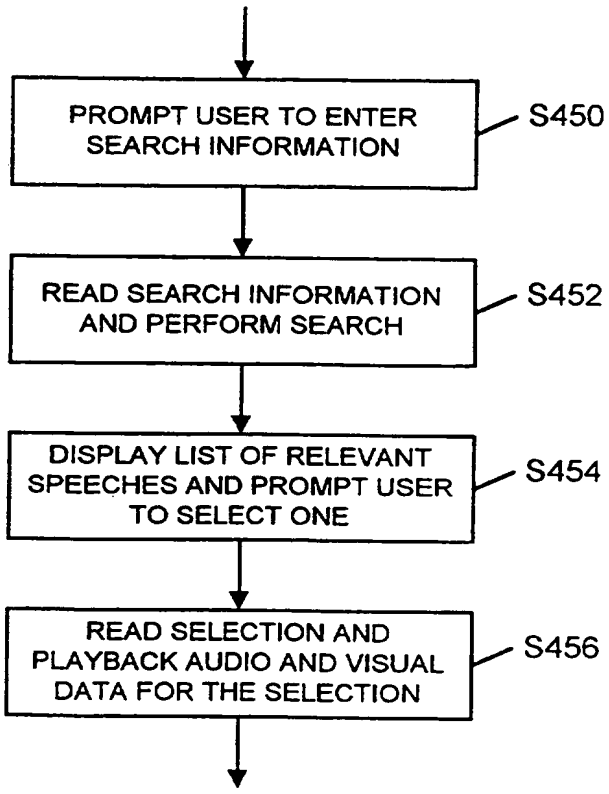

PROMPT USER TO ENTER SEARCH INFORMATION
S450
READ SEARCH INFORMATION AND PERFORM SEARCH
S452
DISPLAY LIST OF RELEVANT SPEECHES AND PROMPT USER TO SELECT ONE
S454
READ SELECTION AND PLAYBACK AUDIO AND VISUAL DATA FOR THE SELECTION
S456

FIG. 38

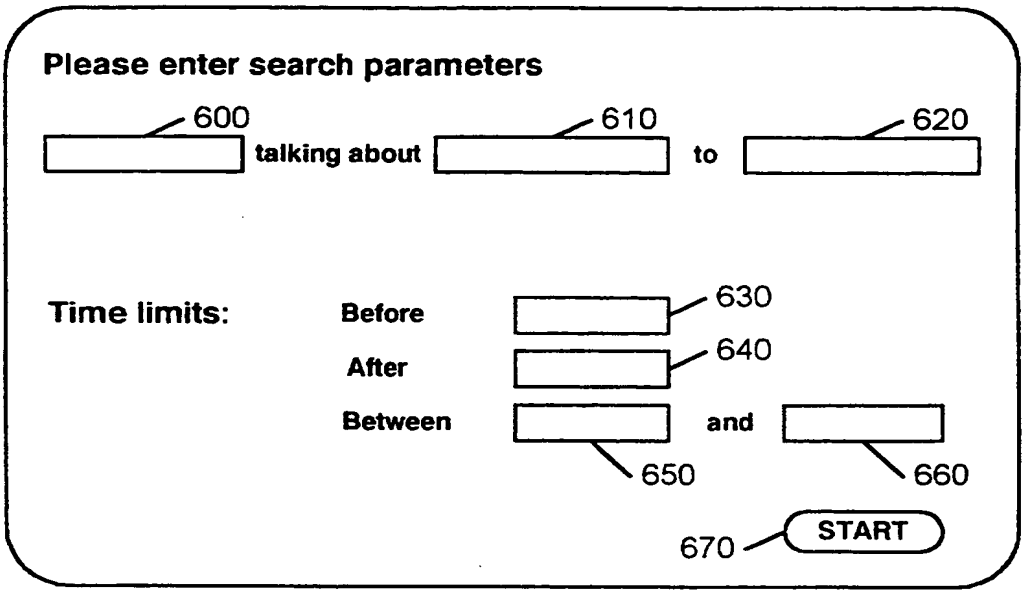


FIG. 39A

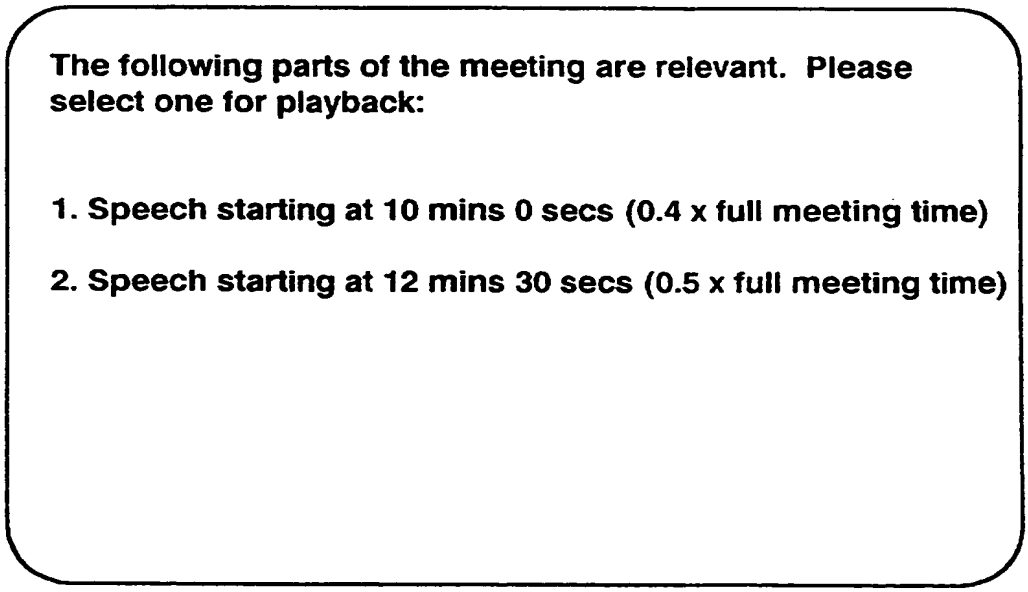


FIG. 39B

FIG. 40

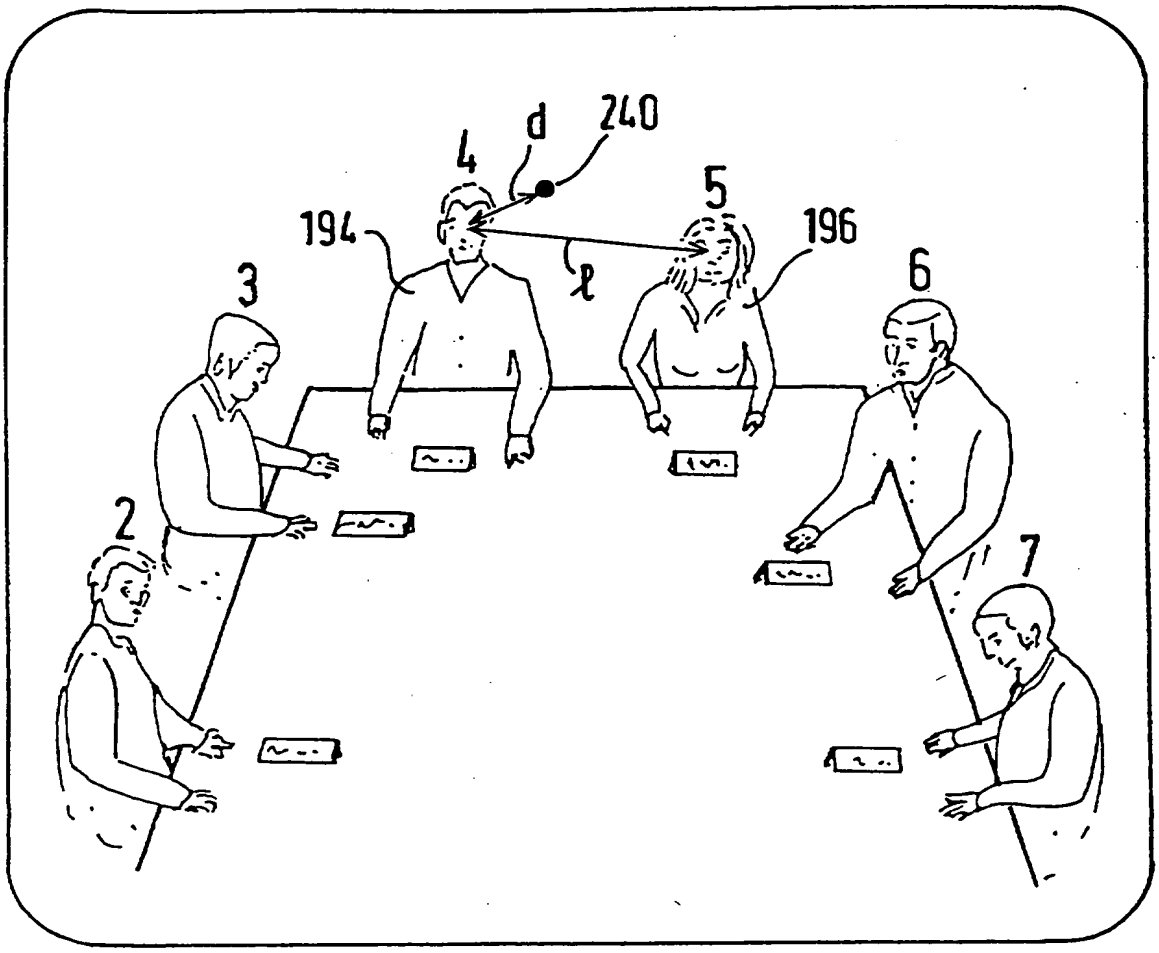


IMAGE PROCESSING APPARATUS

The present invention relates to the field of information archiving, and, more particularly, to the storage of three-dimensional object data.

Many databases exist for the storage of data such as image data etc. However, the existing databases suffer from the problem that it is not possible to store three-dimensional object data therein and/or that the ways in which the database can be interrogated to retrieve information are limited.

It is an object of the present invention to provide a database for the archiving of three-dimensional object data which facilitates improved information retrieval.

According to the present invention, there is provided an apparatus or method in which data for three-dimensional objects is archived together with gaze information to facilitate information retrieval.

The present invention also provides an apparatus or method in which data for three-dimensional objects is stored together with information defining where or at what the objects are looking.

The present invention further provides an apparatus or method for archiving data for three-dimensional objects, in which data defining movements of the three-dimensional objects (or, alternatively data defining the three-dimensional objects in different positions) is stored in association with archive data defining at what (for example a person, animal or object) each three-dimensional object is looking at different times.

Such a system facilitates searching of the stored information to identify archive data for the three-dimensional objects using a query such as "find each part in which Simon is looking at Mike" (Simon and Mike being specific ones of the three-dimensional objects) or such as "find each part in which Allan is looking at mountains".

The present invention further provides an apparatus or method for generating archive data in such a system, and in addition, an apparatus or method for searching stored data in such a system.

The present invention further provides instructions, both in signal and recorded form, for configuring a programmable processing apparatus to become arranged as an apparatus, or to become operable to perform a method,

in such a system.

Embodiments of the invention will now be described, by way of example only, with reference to the accompanying drawings in which:

Figure 1 schematically shows a plurality of user stations interconnected to carry out a video conference in an embodiment of the invention;

Figure 2A shows a user station and a user, Figure 2B shows the headset and body markers worn by the user, and Figure 2C shows the components of the headset worn by the user;

Figure 3 is a block diagram showing an example of notional functional components within the computer processing apparatus at each user station;

Figure 4 shows the steps performed to carry out a video conference;

Figure 5 shows the processing operations performed at step S4 in Figure 4;

Figure 6 shows an example seating plan defined at

step S24 in Figure 5;

Figure 7 shows the processing operations performed at step S6 in Figure 4;

Figure 8 shows the processing operations performed at step S62 in Figure 7;

Figure 9 shows the processing operations performed at step S100 in Figure 8;

Figure 10 shows the processing operations performed at step S130 in Figure 9;

Figure 11 shows the processing operations performed at step S146 and step S150 in Figure 10;

Figure 12 shows the processing operations performed at step S132 in Figure 9;

Figure 13 illustrates the offset angle θ between the plane of the user's head and the plane of his headset calculated at step S64 in Figure 7;

Figure 14 shows the processing operations performed at step S64 in Figure 7;

Figure 15 shows the processing operations performed at step S234 in Figure 14;

Figure 16 illustrates the line projection and mid-point calculation performed at step S252 and step S254 in Figure 15;

Figure 17 shows the processing operations performed at step S66 in Figure 7;

Figure 18 shows the processing operations performed at step S274 in Figure 17;

Figure 19 shows the processing operations performed at step S276 in Figure 17;

Figure 20 shows the processing operations performed at step S324 in Figure 19;

Figure 21 illustrates the angle calculation performed at step S346 in Figure 20;

Figure 22 illustrates the standard coordinate system set up at step S278 in Figure 17;

Figures 23A, 23B, 23C, 23D and 23E show examples of

avatar positions at conference room tables;

Figure 24 shows a piece-wise linear function relating horizontal screen position to view parameter, which is stored at step S72 in Figure 7;

Figure 25 shows the processing operations performed at step S8 in Figure 4;

Figure 26 shows the processing operations performed at step S370 in Figure 25;

Figures 27A, 27B and 27C illustrate the calculation at step S394 in Figure 26 of the point at which the user is looking by projecting a line from the plane of the user's head and determining the intersection of the line with the display screen;

Figure 28 shows the processing operations performed in each of steps S374-1 to S374-6 in Figure 25;

Figures 29A, 29B and 29C illustrate how the position of an avatar's head is changed in dependence upon changes of the corresponding participant's head in real-life at step S430 in Figure 28;

Figure 30 shows the processing operations performed at step S376 in Figure 25;

Figure 31 illustrates examples of markers displayed in images at steps S454 and S456 in Figure 30;

Figure 32 shows the processing operations performed at step S378 in Figure 25;

Figure 33 shows the processing operations performed at step S380 in Figure 25;

Figure 34 shows the processing operations performed at step S400 in Figure 33;

Figure 35 illustrates the three-dimensional computer model of the conference, including the participants' avatars, which is updated at step S400 in Figure 33;

Figures 36A and 36B schematically illustrate the storage of information in the meeting archive database;

Figures 37A, 37B and 37C show examples of viewing histograms stored in the meeting archive database;

Figure 38 shows the processing operations performed at

step S10 in Figure 4;

Figure 39A shows the information displayed to a user at step S450 in Figure 37;

Figure 39B shows an example of information displayed to a user at step S454 in Figure 37; and

Figure 40 schematically illustrates an alternative method for quantising the viewing parameter of a participant at step S404 in Figure 33.

Although an embodiment will now be described relating to a virtual meeting, the invention is not limited to this application, and instead can be applied to any application in which three-dimensional object data is to be archived.

Referring to Figure 1, in this embodiment, a plurality of user stations 2, 4, 6, 8, 10, 12, 14 are connected via a communication path 20, such as the Internet, wide area network (WAN), etc.

As will be described below, each user station 2, 4, 6, 8, 10, 12, 14 comprises apparatus to facilitate a desktop video conference between the users at the user stations.

Figures 2A, 2B and 2C show the components of each user station 2, 4, 6, 8, 10, 12, 14 in this embodiment.

Referring to Figure 2A, a user station comprises a conventional personal computer (PC) 24, two video cameras 26, 28 and a pair of stereo headphones 30.

PC 24 comprises a unit 32 containing, in a conventional manner, one or more processors, memory, and sound card etc, together with a display device 34, and user input devices, which, in this embodiment, comprise a keyboard 36 and mouse 38.

PC 24 is programmed to operate in accordance with programming instructions input for example as data stored on a data storage medium, such as disk 40, and/or as a signal input to PC 24 over a datalink (not shown) such as the Internet, and/or entered by a user via keyboard 36.

PC 24 is connected to the Internet 20 via a connection (not shown) enabling it to transmit data to, and receive data from, the other user stations.

Video cameras 26 and 28 are of conventional design, and are provided to record video images of user 44 using a

charge coupled device (CCD). As will be described below, image data recorded by cameras 26 and 28 is processed by PC 24 to generate data defining the movements of user 44, and this data is then transmitted to the other user stations. Each user station stores a three-dimensional computer model of the video conference containing an avatar for each participant, and each avatar is animated in response to the data received from the user station of the corresponding participant.

In the example shown in Figure 2A, cameras 26 and 28 are positioned on top of monitor 34, but can, however, be positioned elsewhere to view user 44.

Referring to Figures 2A and 2B, a plurality of coloured markers 70, 72 are provided to be attached to the clothing of user 44. The markers each have a different colour, and, as will be explained later, are used to determine the position of the user's torso and arms during the video conference. The markers 70 are provided on elasticated bands to be worn around the user's wrists, elbows and shoulders. A plurality of markers 70 are provided on each elasticated band so that at least one marker will be visible for each position and orientation of the user's arms. The markers 72 are provided with a suitable adhesive so that they can be removably attached

to the torso of user 44, for example along a central line, as shown in Figure 2B, such as at the positions of buttons on the user's clothes.

Referring to Figure 2C, headset 30 comprises earphones 48, 50 and a microphone 52 provided on a headband 54 in a conventional manner. In addition, light emitting diodes (LEDs) 56, 58, 60, 62 and 64 are also provided on headband 54. Each of the LEDs 56, 58, 60, 62 and 64 has a different colour, and, in use, is continuously illuminated. As will be explained later, the LEDs are used to determine the position of the user's head during the video conference.

LED 56 is mounted so that it is central with respect to earphone 48 and LED 64 is mounted so that it is central with respect to earphone 50. The distance "a" between LED 56 and the inner surface of earphone 48 and between LED 64 and the inner surface of earphone 50 is pre-stored in PC 24 for use in processing to be performed during the video conference, as will be described below. LEDs 58 and 62 are slidably mounted on headband 54 so that their positions can be individually changed by user 44. LED 60 is mounted on a member 66 so that it protrudes above the top of headband 54. In this way, when mounted on the head of user 44, LED 60 is held clear of the user's hair.

Each of the LEDs 56, 58, 60, 62 and 64 is mounted centrally with respect to the width of headband 54, so that the LEDs lie in a plane defined by the headband 54.

Signals from microphone 52 and signals to headphones 48, 50 are carried to and from PC 24 via wires in cable 68. Power to LEDs 56, 58, 60, 62 and 64 is also carried by wires in cable 68.

Figure 3 schematically shows the functional units into which the components of PC 24 effectively become configured when programmed by programming instructions. The units and interconnections shown in Figure 3 are notional and are shown for illustration purposes only to assist understanding; they do not necessarily represent the exact units and connections into which the processor, memory, etc of PC 24 become configured.

Referring to Figure 3, central controller 100 processes inputs from user input devices such as keyboard 36 and mouse 38, and also provides control and processing for a number of the other functional units. Memory 102 is provided for use by central controller 100.

Image data processor 104 receives frames of image data recorded by video cameras 26 and 28. The operation of

cameras 26 and 28 is synchronised so that images taken by the cameras at the same time can be processed by image data processor 104. Image data processor 104 processes synchronous frames of image data (one from camera 26 and one from camera 28) to generate data defining (i) image pixel data for the user's face, (ii) the 3D coordinates of each of the markers 70 and 72 on the user's arms and torso, and (iii) a view parameter which, as will be explained further below, defines the direction in which the user is looking. Memory 106 is provided for use by image data processor 104.

The data output by image data processor 104 and the sound from microphone 52 is encoded by MPEG 4 encoder 108. In addition, the sound from microphone 52 is processed by voice recognition processor 107 which operates in accordance with a conventional voice recognition programme, such as "Dragon Dictate" or IBM "ViaVoice", which has previously been trained in a conventional manner to recognise the user's voice. Voice recognition processor 107 generates a stream of text corresponding to the words spoken by the user. The text stream from voice recognition processor 107 is also encoded by MPEG 4 encoder 108. The encoded MPEG 4 bitstream from MPEG 4 encoder 108 is output to the other user stations via input/output interface 110.

Corresponding MPEG 4 bitstreams are received from each of the other user stations and input via input/output interface 110. Each of the bitstreams (bitstream 1, bitstream 2 .... bitstream "n") is decoded by MPEG 4 decoder 112 together with the user's own MPEG 4 bitstream produced by MPEG 4 encoder 108, which is passed to MPEG 4 decoder 112 via the input/output interface 110.

Three-dimensional avatars (computer models) of each of the other participants in the video conference and a three-dimensional computer model of the conference room are stored in avatar and 3D conference model store 114.

In response to the information in the MPEG 4 bitstreams from the other participants, model processor 116 animates the avatars so that the movements of each avatar mimic the movements of the corresponding participant in the video conference.

Image renderer 118 renders an image of the 3D model of the conference room and the avatars, and the resulting pixel data is written to frame buffer 120 and displayed on monitor 34 at a video rate. In this way, images of the avatars and 3D conference model are displayed to the user, and the images show the movement of each avatar corresponding to the movements of the participants in

real-life.

Sound data from the MPEG 4 bitstreams received from the other participants is processed by sound generator 122 together with information from image data processor 104 defining the current position and orientation of the head of user 44, to generate signals which are output to earphones 48 and 50 in order to generate sound to user 44. In addition, signals from microphone 52 are processed by sound generator 22 so that sound from the user's own microphone 52 is heard by the user via his headphones 48 and 50.

Archive processor 124, meeting archive database 126 and text searcher 128 are provided to maintain an archive database of the video conference and to facilitate searching of the database to find and replay the required part of the video conference, as will be described in further detail below.

Figure 4 shows, at a top level, the processing operations carried out to conduct a video conference between the participants at user stations 2, 4, 6, 8, 10, 12 and 14.

Referring to Figure 4, at step S2, suitable connections between each of the user stations 2, 4, 6, 8, 10, 12, 14

are established in a conventional manner.

At step S4, processing operations are performed to set up the video conference. These operations are performed by one of the user stations, previously designated as the conference coordinator.

Figure 5 shows the processing operations performed at step S4 to set up the conference.

Referring to Figure 5, at step S20, the conference coordinator requests the name of each participant, and stores the replies when they are received.

At step S22, the conference coordinator requests the avatar of each participant, and stores the avatars when they are received. Each avatar comprises a three-dimensional computer model of the participant, and may be provided by prior laser scanning of the participant in a conventional manner, or in other conventional ways, for example as described in University of Surrey Technical Report CVSSP - hilton98a, University of Surrey, Guildford, UK.

At step S24, the conference coordinator defines a seating plan for the participants taking part in the video

conference. In this embodiment, this step comprises assigning a number to each participant (including the conference coordinator) and defining the order of the participants around a circular conference room table, for example as shown in Figure 6.

At step S26, the conference room coordinator selects whether a circular or rectangular conference room table is to be used for the video conference.

At step S28, the conference coordinator sends data via Internet 20 defining each of the avatars received at step S22 (including his own), the participant numbers and seating plan defined at step S24, the table shape selected at step S26, and the participants names received at step S20 (including his own) to each of the other participants in the video conference.

Referring again to Figure 4, at step S6, processing operations are performed to calibrate each user station 2, 4, 6, 8, 10, 12, 14 (including the user station of the conference coordinator).

Figure 7 shows the processing operations performed at step S6 to calibrate one of the user stations. These processing operations are performed at every user

station.

Referring to Figure 7, at step S40, the data transmitted by the conference coordinator at step S28 (Figure 5) is received and stored. The three-dimensional avatar model of each participant is stored in its own local reference system in avatar and 3D conference model store 114, and also in its own further local reference system in meeting archive database 126. The other data received from the conference coordinator is stored for example in memory 102 for subsequent use.

At step S42, central controller 100 requests user 44 to input information about the cameras 26, 28. Central controller 100 does this by displaying a message on monitor 34 requesting the user to input for each camera the focal length of the lens in millimetres and the size of the imaging charge couple device (CCD) within the camera. This may be done by displaying on monitor 34 a list of conventional cameras, for which the desired information is pre-stored in memory 102, and from which user 44 can select the camera used, or by the user inputting the information directly. At step S44, the camera parameters input by the user are stored, for example in memory 102 for future use.

At step S46, central controller 100 displays a message on monitor 34 requesting user 44 to input the width in millimetres of the screen of monitor 34, and at step S48, the width which is input by the user is stored, for example in memory 102, for future use.

At step S49, central controller 100 displays a message on monitor 34 instructing the user to wear the headset 30 and body markers 70, 72, as previously described with reference to Figures 2A, 2B and 2C. When the user has completed this step, he inputs a signal to central controller 100 using keyboard 36. Power is then supplied to headset 30 when it is worn by user 44 so that each of the LEDs 56, 58, 60, 62 and 64 are continuously illuminated.

At step S50, central controller 100 displays a message on monitor 34 instructing the user to position the movable LEDs 58, 62, on headset 30 so that the LEDs align with the user's eyes. When the user has slid LEDs 58 and 62 on headband 54 so that they align with his eyes, he inputs a signal to central controller 100 using keyboard 36.

At step S52, central controller 100 displays a message on monitor 34 instructing the user to position cameras 26

and 28 so that both cameras have a field of view which covers the user's position in front of PC 24. When the user has positioned the cameras, he inputs a signal to central controller 100 using keyboard 36.

At step S54, central controller 100 displays a message on monitor 34 instructing the user to move backwards, forwards, and to each side over the full range of distances that the user is likely to move during the video conference. At step S56, as the user moves, frames of image data are recorded by cameras 26 and 28 and displayed on monitor 34, so that the user can check whether he is visible to each camera at all positions.

At step S58, central controller 100 displays a message on monitor 34 asking the user whether it is necessary to adjust the positions of the cameras so that the user is visible throughout the full range of his likely movements. If the user inputs a signal using keyboard 36 indicating that camera adjustment is necessary, steps S52 to S58 are repeated until the cameras are correctly positioned. On the other hand, if the user inputs a signal indicating that the cameras are correctly positioned, then processing proceeds to step S60.

At step S60, central controller 100 processes the data

defining the avatar of user 44 to determine the user's head ratio, that is, the ratio of the width of the user's head (defined by the distance between the user's ears) and the length of the user's head (defined by the distance between the top of the user's head and the top of his neck), and also the width of the user's head in real-life (which can be determined since the scale of the avatar is known). The head ratio and real-life width are stored, for example in memory 106 for subsequent use by the image data processor 104.

At step S62, central controller 100 and image data processor 104 use the frames of image data previously recorded at step S56 (after the cameras 26 and 28 had been positioned for the final time) to determine the camera transformation model to be used during the video conference. The camera transformation model defines the relationship between the image plane (that is, the plane of the CCD) of camera 26 and the image plane of camera 28 which will be used to reconstruct the three-dimensional positions of the headset LEDs 56, 58, 60, 62, 64 and the body markers 70, 72 using images of these LEDs and markers recorded by the cameras 26 and 28.

Figure 8 shows the processing operations performed by central controller 100 and image data processor 104 at

step S62 to determine the camera transformation model.

Referring to Figure 8, at step S90, the frames of image data recorded at step S56 are processed to identify the pair of synchronous images (that is, the image from camera 26 and the image from camera 28 recorded at the same time) which show the most left position, the pair which show the most right position, the pair which show the most forward position, and the pair which show the most backward position to which the user moved. In this embodiment, step S90 is performed by displaying the sequence of images recorded by one of the cameras at step S56, and instructing the user to input a signal, for example via keyboard 36 or mouse 38, when the image for each of the extreme positions is displayed. As noted above, these positions represent the extents of the user's likely movement during the video conference. Images for the most left position and most right position are identified and considered in subsequent processing to determine the camera transformation model since each of the cameras 26 and 28 is positioned at an angle to the user, and so movement of the user to the right or left increases or decreases the distance of the user from each of the cameras.

At step S92, the image data for each of the four pairs

of images identified at step S90 (that is, the pair of images for the most left position, the pair of images for the most right position, the pair of images for the most forward position and the pair of images for the most backward position) is processed to identify the positions of the LEDs 56, 58, 60, 62, 64 and coloured body markers 70, 72 which are visible in each image of the pair and to match each of the identified points between the images in the pair. In this step, since each LED and each body marker has a unique colour, the pixel data for each image in a synchronised pair is processed to identify those pixels having one of the predetermined colours by examining the RGB values of the pixels. Each group of pixels having one of the predetermined colours is then processed using a convolution mask to find the coordinates within the image as a whole of the centre of the group of pixels. This is performed in a conventional manner, for example as described in "Affine Analysis of Image Sequences" by L.S. Shapiro, Cambridge University Press, 1995, ISBN 0-521-55063-7, pages 16-23. The matching of points between images is done by identifying the point in each image which has the same colour (if a marker or LED is visible to only one of the cameras 26 or 28, and hence appears in only one image, then no matched pair of points will be identified for this LED or marker).

At step S94, the coordinates of the matched points identified at step S92 are normalised. Up to this point, the coordinates of the points are defined in terms of the number of pixels across and down an image from the top left hand corner of the image. At step S94, the camera focal length and image plane size previously stored at step S44 are used to convert the coordinates of the points from pixels to a coordinate system in millimetres having an origin at the camera optical centre. The millimetre coordinates are related to the pixel coordinates as follows:

$$x^* = h \times (x - C_x) \quad \ldots.(1)$$

$$y^* = -v \times (y - C_y) \quad \ldots.(2)$$

where (x^*,y^*) are the millimetre coordinates, (x,y) are the pixel coordinates, (C_x,C_y) is the centre of the image (in pixels), which is defined as half of the number of pixels in the horizontal and vertical directions, and "h" and "v" are the horizontal and vertical distances between adjacent pixels (in mm).

At step S96, a set is formed of all the matched points identified at step S92. This set therefore contains points for all four pairs of images. Of course, the

number of points in the combined set from each pair of images may be different, depending upon which LEDs and body markers are visible in the images. However the large number of body markers and LEDs ensures that at least seven markers or LEDs will be visible in each image, giving a minimum of 4 × 7 = 28 pairs of matched points in the combined set.

At step S98, a measurement matrix, M, is set up as follows for the points in the combined set created at step S96:

$$M = \begin{pmatrix} x_1x_1' & -y_1x_1' & x_1' & -x_1y_1' & y_1y_1' & -y_1' & x_1 & -y_1 & 1 \\ x_2x_2' & -y_2x_2' & x_2' & -x_2y_2' & y_2y_2' & -y_2' & x_2 & -y_2 & 1 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ x_kx_k' & -y_kx_k' & x_k' & -x_ky_k' & y_ky_k' & -y_k' & x_k & -y_k & 1 \end{pmatrix} \quad \ldots(3)$$

where (x,y) are the pixel coordinates of the point in the first image of a pair, (x',y') are the pixel coordinates of the corresponding (matched) point in the second image of the pair, and the numbers 1 to k indicate to which pair of points the coordinates correspond (there being k pairs of points in total).

At step S100, the most accurate camera transformation for

the matched points in the combined set is calculated. By calculating this transformation using the combined set of points created at step S96, the transformation is calculated using points matched in a pair of images representing the user's most left position, a pair of images representing the user's most right position, a pair of images representing the user's most forward position, and a pair of images representing the user's most backward position. Accordingly, the calculated transformation will be valid over the user's entire workspace.

Figure 9 shows the processing operations performed at step S100 to calculate the most accurate camera transformation.

Referring to Figure 9, at step S130, a perspective transformation is calculated, tested and stored.

Figure 10 shows the processing operations performed at step S130.

Referring to Figure 10, at step S140, the next seven pairs of matched points in the combined set created at step S96 are selected (this being the first seven pairs the first time step S140 is performed).

At step S142, the selected seven pairs of points and the measurement matrix set at step S98 are used to calculate the fundamental matrix, F, representing the geometrical relationship between the cameras, F being a three by three matrix satisfying the following equation:

$$(x' \; y' \; 1) F \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0 \qquad \ldots\ldots(4)$$

where (x,y,1) are the homogeneous pixel coordinates of any of the seven selected points in the first image of the pair, and (x',y',1) are the corresponding homogeneous pixel coordinates in the second image of the pair.

The fundamental matrix is calculated in a conventional manner, for example using the technique disclosed in "Robust Detection of Degenerate Configurations Whilst Estimating the Fundamental Matrix" by P.H.S. Torr, A. Zisserman and S. Maybank, Oxford University Technical Report 2090/96.

It is possible to select more than seven pairs of matched points at step S140 and to use these to calculate the fundamental matrix at step S142. However, seven pairs of points are used in this embodiment, since this has been shown empirically to produce satisfactory results, and also represents the minimum number of pairs needed

to calculate the parameters of the fundamental matrix, reducing processing requirements.

At step S144, the fundamental matrix, F, is converted into a physical fundamental matrix, F_{phys}, using the camera data stored at step S44 (Figure 7). This is again performed in a conventional manner, for example as described in "Motion and Structure from Two Perspective Views: Algorithms, Error Analysis and Error Estimation" by J. Weng, T.S. Huang and N. Ahuja, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, No. 5, May 1989, pages 451-476, and as summarised below.

First the essential matrix, E, which satisfies the following equation is calculated:

$$(x^{*\prime}\ y^{*\prime}\ f)\,E\begin{pmatrix} x^{*} \\ y^{*} \\ f \end{pmatrix} = 0 \qquad \ldots\ldots(5)$$

where (x*, y*, f) are the coordinates of any of the selected seven points in the first image in a millimetre coordinate system whose origin is at the centre of the image, the z coordinate having being normalised to correspond to the focal length, f, of the camera, and (x*', y*', f) are the corresponding coordinates of the matched point in the second image of the pair. The fundamental matrix, F, is converted into the essential

matrix, E, using the following equations:

$$A = \begin{pmatrix} 1/h & 0 & c_x/f \\ 0 & 1/v & -c_y/f \\ 0 & 0 & 1/f \end{pmatrix} \quad \ldots.(6)$$

$$M = A^T F A \quad \ldots.(7)$$

$$E = \sqrt{\frac{2}{tr(M^T M)}} \times M \quad \ldots.(8)$$

where the camera parameters "h", "v", "c_x", "c_y" and "f" are as defined previously, the symbol T denotes the matrix transpose, and the symbol "tr" denotes the matrix trace.

The calculated essential matrix, E, is then converted into a physical essential matrix, "E_{phys}", by finding the closest matrix to E which is decomposable directly into a translation vector (of unit length) and rotation matrix (this closest matrix being E_{phys}).

Finally, the physical essential matrix is converted into a physical fundamental matrix, using the equation:

$$F_{phys} = A^{-1T} E_{phys} A^{-1} \quad \ldots.(9)$$

where the symbol "-1" denotes the matrix inverse.

Each of the physical essential matrix, E_{phys}, and the physical fundamental matrix, F_{phys} is a "physically realisable matrix", that is, it is directly decomposable into a rotation matrix and translation vector.

The physical fundamental matrix, F_{phys}, defines a curved surface in a four-dimensional space, represented by the coordinates (x, y, x', y') which are known as "concatenated image coordinates". The curved surface is given by Equation (4) above, which defines a 3D quadric in the 4D space of concatenated image coordinates.

At step S146, the calculated physical fundamental matrix is tested against each pair of points that were used to calculate the fundamental matrix at step S142. This is done by calculating an approximation to the 4D Euclidean distance (in the concatenated image coordinates) of the 4D point representing each pair of points from the surface representing the physical fundamental matrix. This distance is known as the "Sampson distance", and is calculated in a conventional manner, for example as described in "Robust Detection of Degenerate Configurations Whilst Estimating the Fundamental Matrix" by P.H.S. Torr, A. Zisserman and S. Maybank, Oxford University Technical Report 2090/96.

Figure 11 shows the processing operations performed at step S146 to test the physical fundamental matrix.

Referring to Figure 11, at step S170, a counter is set to zero. At step S172, the tangent plane of the surface representing the physical fundamental matrix at the four-dimensional point defined by the coordinates of the next pair of points in the seven pairs of points (the two coordinates defining each point in the pair being used to define a single point in the four-dimensional space of the concatenated image coordinates) is calculated. Step S172 effectively comprises shifting the surface to touch the point defined by the coordinates of the pair of points, and calculating the tangent plane at that point. This is performed in a conventional manner, for example as described in "Robust Detection of Degenerate Configurations Whilst Estimating the Fundamental Matrix" by P.H.S. Torr, A. Zisserman and S. Maybank, Oxford University Technical Report 2090/96.

At step S174, the normal to the tangent plane determined at step S172 is calculated, and, at step S176, the distance along the normal from the point in the 4D space defined by the coordinates of the pair of matched points to the surface representing the physical fundamental matrix (the "Sampson distance") is calculated.

At step S178, the calculated distance is compared with a threshold which, in this embodiment, is set at 1.0 pixels. If the distance is less than the threshold, then the point lies sufficiently close to the surface, and the physical fundamental matrix is considered to accurately represent the relative positions of the cameras 26 and 28 for the particular pair of matched points being considered. Accordingly, if the distance is less than the threshold, at step S180, the counter which was initially set to zero at step S170 is incremented, the points are stored, and the distance calculated at step S176 is stored.

At step S182, it is determined whether there is another pair of points in the seven pairs of points used to calculate the fundamental matrix, and steps S172 to S182 are repeated until all such points have been processed as described above.

Referring again to Figure 10, at step S148, it is determined whether the physical fundamental matrix calculated at step S144 is sufficiently accurate to justify further processing to test it against all of the pairs of matched points in the combined set. In this embodiment, step S148 is performed by determining whether the counter value set at step S180 (indicating the number

of pairs of points which have a distance less than the threshold tested at step S178, and hence are considered to be consistent with the physical fundamental matrix) is equal to 7. That is, it is determined whether the physical fundamental matrix is consistent with all of the points used to calculate the fundamental matrix from which the physical fundamental matrix was derived. If the counter is less than 7, the physical fundamental matrix is not tested further, and processing proceeds to step S152. On the other hand, if the counter value is equal to 7, at step S150, the physical fundamental matrix is tested against each other pair of matched points. This is performed in the same way as step S146 described above, with the following exceptions: (i) at step S170, the counter is set to 7 to reflect the seven pairs of points already tested at step S146 and determined to be consistent with the physical fundamental matrix, and (ii) the total error for all points stored at step S180 (including those stored during processing at step S146) is calculated, using the following equation:

$$Total\ error = \frac{\sqrt{\sum \frac{e_i^2}{p}}}{e_{th}} \qquad \ldots.(10)$$

where e_i is the distance for the "i"th pair of matched

points between the 4D point represented by their coordinates and the surface representing the physical fundamental matrix calculated at step S176, this value being squared so that it is unsigned (thereby ensuring that the side of the surface representing the physical fundamental matrix on which the point lies does not affect the result), p is the total number of points stored at step S180, and e_{th} is the distance threshold used in the comparison at step S178.

The effect of step S150 is to determine whether the physical fundamental matrix calculated at step S144 is accurate for each pair of matched points in the combined set, with the value of the counter at the end (step S180) indicating the total number of the points for which the calculated matrix is sufficiently accurate.

At step S152, it is determined whether the physical fundamental matrix tested at step S150 is more accurate than any previously calculated using the perspective calculation technique. This is done by comparing the counter value stored at step S180 in Figure 11 for the last-calculated physical fundamental matrix (this value representing the number of points for which the physical fundamental matrix is an accurate camera solution) with the corresponding counter value stored for the most

accurate physical fundamental matrix previously calculated. The matrix with the highest number of points (counter value) is taken to be the most accurate. If the number of points is the same for two matrices, the total error for each matrix (calculated as described above) is compared, and the most accurate matrix is taken to be the one with the lowest error. If it is determined at step S152 that the physical fundamental matrix is more accurate than the currently stored one, then, at step S154 the previous one is discarded, and the new one is stored together with the number of points (counter value) stored at step S180 in Figure 11, the points themselves, and the total error calculated for the matrix.

At step S156, it is determined whether there is another pair of matched points which has not yet been considered, such that there is another unique set of seven pairs of matched points in the combined set to be processed. Steps S140 to S156 are repeated until each unique set of seven pairs of matched points has been processed in the manner described above.

Referring again to Figure 9, at step S132, an affine relationship for the matched points in the combined set is calculated, tested and stored.

Figure 12 shows the processing operations performed at step S132.

Referring to Figure 12, at step S200, the next four pairs of matched points are selected for processing (this being the first four pairs the first time step S200 is performed).

When performing the perspective calculations (step S130 in Figure 9), it is possible to calculate all of the components of the fundamental matrix, F. However, when the relationship between the cameras is an affine relationship, it is possible to calculate only four independent components of the fundamental matrix, these four independent components defining what is commonly known as an "affine" fundamental matrix.

Accordingly, at step S202, the four pairs of points selected at step S200 and the measurement matrix set at step S96 are used to calculate four independent components of the fundamental matrix (giving the "affine" fundamental matrix) using a technique such as that described in "Affine Analysis of Image Sequences" by L.S. Shapiro, Section 5, Cambridge University Press 1995, ISBN 0-521-55063-7. It is possible to select more than four pairs of points at step S200 and to use these to

calculate the affine fundamental matrix at step S202. However, in the present embodiment, only four pairs are selected since this has been shown empirically to produce satisfactory results, and also represents the minimum number required to calculate the components of the affine fundamental matrix, reducing processing requirements.

At step S204, the affine fundamental matrix is tested against each pair of matched points in the combined set using a technique such as that described in "Affine Analysis of Image Sequences" by L.S. Shapiro, Section 5, Cambridge University Press, 1995, ISBN 0-521-55063-7. The affine fundamental matrix represents a flat surface (hyperplane) in four-dimensional, concatenated image space, and this test comprises determining the distance between a point in the four-dimensional space defined by the coordinates of a pair of matched points and the flat surface representing the affine fundamental matrix. As with the tests performed during the perspective calculations at step S146 and S150 (Figure 10), the test performed at step S204 generates a value for the number of pairs of points for which the affine fundamental matrix represents a sufficiently accurate solution to the camera transformation and a total error value for these points.

At step S206, it is determined whether the affine fundamental matrix calculated at step S202 and tested at step S204 is more accurate than any previously calculated. This is done by comparing the number of points for which the matrix represents an accurate solution with the number of points for the most accurate affine fundamental matrix previously calculated. The matrix with the highest number of points is the most accurate. If the number of points is the same, the matrix with the lowest error is the most accurate. If the affine fundamental matrix is more accurate than any previously calculated, then at step S208, it is stored together with the points for which it represents a sufficiently accurate solution, the total number of these points and the matrix total error.

At step S210, it is determined whether there is another pair of matched points to be considered, such that there exists another unique set of four pairs of matched points in the combined set to be processed. Steps S200 to S210 are repeated until each unique set of four pairs of matched points are processed in the manner described above.

Referring again to Figure 9, at step S134, the most accurate transformation is selected from the perspective

transformation calculated at step S130 and the affine transformation calculated at step S132. This step is performed by comparing the number of points which are consistent with the most accurate perspective transformation (stored at step S154) with the number of points which are consistent with the most accurate affine transformation (stored at step S208), and selecting the transformation which has the highest number of consistent points (or the transformation having the lowest matrix total error if the number of consistent points is the same for both transformations).

At step S104, it is determined whether the affine transformation is the most accurate camera transformation. If it is determined at step S104 that the affine transformation is not the most accurate transformation, then, at step S106, the perspective transformation which was determined at step S100 is selected for use during the video conference.

At step S108, the physical fundamental matrix for the perspective transformation is converted to a camera rotation matrix and translation vector. This conversion is performed in a conventional manner, for example as described in the above-referenced "Motion and Structure from Two Perspective Views: Algorithms, Error Analysis

and Error Estimation" by J. Weng, T.S. Huang and N. Ahuja, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 5, May 1989, pages 451-476.

In the processing described above with respect to Figure 10, a fundamental matrix is calculated (steps S142) and converted to a physical fundamental matrix (step S144) for testing against the matched points (steps S146 and S150). This has the advantage that, although additional processing is required to convert the fundamental matrix to a physical fundamental matrix, the physical fundamental matrix ultimately converted at step S108 has itself been tested. If the fundamental matrix was tested, this would then have to be converted to a physical fundamental matrix which would not, itself, have been tested.

On the other hand, if it is determined at step S104, that the affine transformation is the most accurate transformation, then, at step S110, the affine transformation is selected for use during the video conference.

At step S112, the affine fundamental matrix is converted into three physical variables describing the camera transformation, namely the magnification, "m", of the

object between images recorded by the cameras, the axis, ϕ, of rotation of the camera, and the cyclotorsion rotation, θ, of the camera. The conversion of the affine fundamental matrix into these physical variables is performed in a conventional manner, for example as described in "Affine Analysis of Image Sequences" by L.S. Shapiro, Cambridge University Press, 1995, ISBN 0-521-55063-7, Section 7.

Referring again to Figure 7, at step S64, the position of the headset LEDs 56, 58, 60, 62 and 64 relative to the head of user 44 is determined. This step is performed since this relative position will depend on how the user has placed the headset 30 on his head. More particularly, as illustrated in Figure 13, the plane 130 in which the headset LEDs lie is determined by the angle at which the user wears the headset 30. Accordingly, the plane 130 of the headset LEDs may be different to the actual plane 132 of the user's head. At step S64, therefore, processing is carried out to determine the angle θ between the plane 130 of the headset LEDs and the actual plane 132 of the user's head.

Figure 14 shows the processing operations performed at step S64.

Referring to Figure 14, at step S230, central controller 100 displays a message on monitor 34 instructing the user 44 to look directly at the camera to his right (that is, camera 28 in this embodiment).

At step S232, a frame of image data is recorded with both camera 26 and camera 28 while the user is looking directly at camera 28.

At step S234, the synchronous frames of image data recorded at step S232 are processed to calculate the 3D positions of the headset LEDs 56, 58, 60, 62 and 64.

Figure 15 shows the processing operations performed at step S324 to calculate the 3D positions of the headset LEDs.

Referring to Figure 15, at step S250, the position of each headset LED 56, 58, 60, 62 and 64 is identified in each of the images recorded at step S232. The identification of the LED positions at step S250 is carried out in the same way as previously described with respect to step S92 (Figure 8).

At step S252, the positions of the next pair of LEDs matched between the pair of images are considered, and

the camera transformation model previously determined at step S62 (Figure 7) is used to calculate the projection of a ray from the position of the LED in the first image through the optical centre of the camera for the first image, and from the position of the matched LED in the second image through the optical centre of the camera for the second image. This is illustrated in Figure 16. Referring to Figure 16, ray 140 is projected from the position of an LED (such as LED 56) in the image 142 recorded by camera 26 through the optical centre of camera 26 (not shown), and ray 144 is projected from the position of the same LED in image 146 recorded by camera 28, through the optical centre of camera 28 (not shown).

Referring again to Figure 15, at step S254, the mid-point 148 (Figure 16) of the line segment which connects, and is perpendicular to, both of the rays projected in step S252 is calculated. The position of this mid-point represents the physical position of the LED in three dimensions.

At step S256, it is determined whether there is another one of the LEDs 56, 58, 60, 62 or 64 to be processed. Steps S252 to S256 are repeated until the three-dimensional coordinates of each of the LEDs has been calculated as described above.

Referring again to Figure 14, at step S236, the plane 130 (Figure 13) in which the three-dimensional positions of the headset LEDs lie is determined, and the angle θ between this plane and the imaging plane of the camera at which the user was looking when the frames of image data were recorded at step S232 is calculated. Since the user was looking directly at the camera to his right when the frames of image data were recorded at step S232, the direction of the imaging plane of the camera to the user's right corresponds to the direction of the plane 132 of the user's head (Figure 13). Accordingly, the angle calculated at step S236 is the angle θ between the plane 130 of the headset LEDs and the plane 132 of the user's head.

Referring again to Figure 7, at step S66, the position of the display screen of monitor 34 is determined and a coordinate system is defined relative to this position.

Figure 17 shows the processing operations performed at step S66.

Referring to Figure 17, at step S270, the central controller 100 displays a message on monitor 34 instructing the user to sit centrally and parallel to the display screen of the monitor 34, and to sit upright with

his torso touching the edge of the desk on which PC 24 stands. At step S272, a further message is displayed instructing the user to turn but not otherwise change the position of, his head, so that the processing in the steps which follow can be carried out on the basis of a constant head position but changing head angle.

At step S274, the direction of the plane of the display screen of monitor 34 is determined. In this embodiment, this is done by determining the direction of a plane parallel to the display screen.

Figure 18 shows the processing operations performed at step S274.

Referring to Figure 18, at step S300, central controller 100 displays a marker in the centre of the display screen of monitor 34, and instructs the user to look directly at the displayed marker.

At step S302, a frame of image data is recorded with both camera 26 and 28 as the user looks at the displayed marker in the centre of the screen of monitor 34.

At step S304, the three-dimensional positions of the coloured markers 72 on the user's torso are determined.

This step is carried out in the same way as step S234 in Figure 14, which was described above with respect to Figures 15 and 16, the only difference being that, since the positions of the coloured markers 72 in each image are determined (rather than the positions of the headset LEDs), rays are projected from the positions of matched markers in each of the synchronised images. Accordingly, these steps will not be described again here.

At step S306, the three-dimensional positions of the user's headset LEDs are calculated. This step is also carried out in the same way as step S234 in Figure 14, described above with respect to Figures 15 and 16.

At step S308, the plane in which the three-dimensional positions of the headset LEDs (determined at step S306) lie is calculated.

At step S310, the direction of the plane determined at step S308 is adjusted by the angle θ determined at step S64 (Figure 7) between the plane of the headset LEDs and the plane of the user's head. The resulting direction is the direction of a plane parallel to the plane of the display screen, since the plane of the user's head will be parallel to the display screen when the user is looking directly at the marker in the centre

of the screen.

Referring again to Figure 17, at step S276, the position in three dimensions of the plane of the display screen of monitor 34 is determined.

Figure 19 shows the processing operations performed at step S276.

Referring to Figure 19, at step S320, central controller 100 displays a marker in the centre of the right edge of the display screen of monitor 34, and displays a message instructing the user to look at the marker.

At step S322, a frame of image data is recorded with both camera 26 and 28 as the user looks at the marker displayed at the edge of the display screen.

At step S324, the angle of the user's head relative to the display screen about a vertical axis is determined.

Figure 20 shows the processing operations performed at step S324.

Referring to Figure 20, at step S340, the three-

dimensional positions of the headset LEDs are calculated. This step is carried out in the same manner as step S234 in Figure 14, and described above with respect to Figures 15 and 16. Accordingly, the processing operations will not be described again here.

At step S342, the plane which passes through the three-dimensional positions of the headset LEDs is determined, and, at step S344, the position of this plane is adjusted by the headset offset angle θ (calculated at step S64 in Figure 7) to give the plane of the user's head.

At step S346, the angle between the direction of the plane of the user's head determined at step S344 and the direction of the plane parallel to the display screen determined at step S274 (Figure 17) is calculated. This calculated angle is the angle of the user's head relative to the plane of the display screen about a vertical axis, and is illustrated in Figure 21 as angle α.

Referring again to Figure 19, at step S326, the three-dimensional position of the display screen is calculated and stored for subsequent use. In this step, the width of the display screen previously input by the user at step S46 and stored at step S48 (Figure 7) is used together with the angle determined at step S324 of the

user's head when looking at a point at the edge of the display screen to calculate the 3D position of the display screen. More particularly, referring to Figure 21, the distance "d" of the plane parallel to the display screen determined at step S274 (Figure 17) is calculated using the angle α and one half of the width "W" of the display screen, thereby determining the three-dimensional position of the plane of the display screen. The extents of the display screen in the horizontal direction are then determined using the width "W".

Referring again to Figure 17, at step S278, a three-dimensional coordinate system and scale is defined relative to the three-dimensional position of the display screen. This coordinate system will be used to define the three-dimensional position of points which are transmitted to the other participants during the video conference. Accordingly, each participant uses the same coordinate system and scale, and therefore transmits coordinates which can be interpreted by the other participants. Referring to Figure 22, in this embodiment, the coordinate system is defined with the origin at the centre of the display screen, the "x" and "y" axes lying in the plane of the display screen in horizontal and vertical directions respectively, and the "z" axis lying in a direction perpendicular to the plane

of the display screen in a direction towards the user. The scale for each axis is predefined (or could, for example, be transmitted to each user station by the conference coordinator).

Also at step S278, the transformation is calculated which maps three-dimensional coordinates calculated using the camera transformation model determined at step S62 to the new, standardised coordinate system and scale. This transformation is calculated in a conventional manner, with scale changes being determined by using the width of the user's head in real-life (determined at step S60 in Figure 7) and the distance "a" between each of LEDs 56 and 64 and the inner surface of the earphones 48, 50 (Figure 2C) to determine the distance between the LEDs 56 and 64 in real-life when the headset 30 is worn by the user, and by using this real-life LED separation to relate the distance between the three-dimensional coordinates of the headset LEDs 56 and 64 calculated using the camera transformation model at step S306 in Figure 18 to the predefined scale of the standard coordinate system.

At step S280, the three-dimensional positions of the body markers 72 previously calculated at step S304 (Figure 18) are transformed into the standard coordinate system

defined at step S278.

At step S282, the three-dimensional positions of the body markers 72 in the standard coordinate system are transmitted to the other participants in the video conference, for subsequent use in positioning the user's avatar in the three-dimensional computer model of the conference room stored in the avatar and 3D conference model store 114 and in the three-dimensional computer model stored in the meeting archive database 126, as will be described below.

Referring again to Figure 7, at step S68, a three-dimensional computer model is set up of the conference room table to be used for the video conference, and is stored in the avatar and 3D conference model store 114. In this embodiment, three-dimensional computer models are pre-stored of a rectangular and semi-circular conference room table, and the appropriate model is selected for use in dependence upon the instructions received from the conference room coordinator at step S40 defining the shape of the conference room table to be used.

In addition, name labels showing the name of each of the participants are placed on the conference room table in the three-dimensional computer model, with the name

displayed on each label being taken from the names of the participants received from the conference coordinator at step S40. In order to determine the positions for the name labels on the conference table, the seating position of each participant is first determined using the seating plan received from the conference coordinator at step S40. Although the conference coordinator defined the seating plan by defining the order of the participants in a circle (step S24 in Figure 5, and Figure 6), at step S68 the positions of the avatars around the conference room table are set so that, when an image of the avatars and conference room table is displayed to the user, the avatars are spread apart across the width of the display screen of monitor 34. In this way, each avatar occupies its own part of the display screen in the horizontal direction and all of the avatars can be seen by the user.

Figures 23A, 23B, 23C, 23D and 23E illustrate how the positions of avatars are set in this embodiment for different numbers of participants in the video conference. Referring to Figures 23A, 23B, 23C, 23D and 23E in general, the avatars are spaced apart evenly around a semi-circle 164 in three dimensions. The diameter of the semi-circle 164 (which is the same irrespective of the number of participants in the video

conference) and the viewing position from which images are rendered for display to the user are chosen so that each avatar occupies a unique position across the display screen and the outermost avatars are close to the edges of the display screen in the horizontal direction. In this embodiment, the avatars are positioned around semi-circle 164 and a viewing position is defined such that the positions at which the avatars appear in an image are shown in the table below.

NUMBER OF AVATARS DISPLAYED	POSITION OF AVATAR IN IMAGE (W = screen width)
2	±0.46W
3	0.00W; ±0.46W
4	±0.20W; ±0.46W
5	0.00W; ±0.20W; ±0.46W
6	±0.12W; ±0.34W; ±0.46W

Table 1

Referring to Figure 23A, when there are three participants in the video conference, the avatars 160 and 162 for the two participants other than the user at the user station being described are positioned behind the same, straight edge of a conference room table at the ends of the semi-circle 164. As set out in the table above, avatar 160 is positioned so that it appears in an image at a distance -0.46W from the centre of the display

screen in a horizontal direction, and avatar 162 is positioned so that it appears at a distance +0.46W from the centre. Name plates 166 and 168 showing the respective names of the participants are placed on the conference room table in front of the avatars facing the viewing position from which images of the conference room table and avatars will be rendered. In this way, the user, when viewing the display, can read the name of each participant.

Figure 23B shows an example in which there are four participants of the video conference and a rectangular conference room table has been selected by the conference organiser. Again, the avatars 170, 172 and 174 for the three participants other than the user at the user station are arranged around the semi-circle 164 with equal spacing. Avatar 170 is positioned so that it appears in an image at a distance -0.46W from the centre of the display screen in a horizontal direction, avatar 172 is positioned so that it appears at the centre of the display screen (in a horizontal direction), and avatar 174 is positioned so that it appears at a distance +0.46W from the centre. A name label 176, 178, 180 is placed on the conference room table facing the viewing position from which images of the conference room table and avatars will be rendered.

Figure 23C shows an example in which there are four participants of the video conference, as in the example of Figure 23B, but the conference coordinator has selected a circular conference room table. In this case, the edge of the model of the conference room table follows the semi-circle 164.

Figure 23D shows an example in which there are seven participants in the video conference, and a rectangular conference room table is specified by the conference coordinator. The avatars 190, 192, 194, 196, 198, 200 for each of the participants other than the user at the user station are equally spaced around semi-circle 164, such that, when an image is rendered, the avatars occupy positions of -0.46W, -0.34W, -0.12W, +0.12W, +0.34W and +0.46W respectively from the centre of the display screen in a horizontal direction. A name label 202, 204, 206, 208, 210, 212 is provided for each participant facing the viewing position from which images will be rendered so that the participants' names are visible in the image displayed on monitor 34 to the user.

The relative positions and orientations of the avatars around the conference room table will be different for the participant at each user station. Referring to the seating plan shown in Figure 6, and assuming that the

user at the user station being described is participant 1, then participant 2 is to the left of the user and participant 7 is to the right of the user. Accordingly, as shown in Figure 23D, the position of avatar 190 for participant 2 is set so that it appears on the left of the image, and the position of avatar 200 for participant 7 is set so that it appears on the right of the image. The positions of avatars 192, 194, 196 and 198 for participants 3, 4, 5 and 6 respectively are arranged between the positions of avatars 190 and 200 in accordance with the order defined in the seating plan.

Similarly, by way of further example, the positions of the avatars would be set at the user station of participant 2 so that the order of the participants from left to right in an image is 3, 4, 5, 6, 7 and 1.

The example shown in Figure 23E corresponds to that shown in Figure 23D, except that a circular conference room table is specified by the conference coordinator.

Referring again to Figure 7, at step S70, a respective transformation is defined for each participant which maps the avatar for the participant from the local coordinate system in which it was stored at step S40 into the three-dimensional computer model of the conference room created

at step S68 so that the avatar appears at the correct position at the conference room table. In this step, the three-dimensional positions of the body markers 72 previously received from each participant (as transmitted at step S282 in Figure 17) when the participant was sitting with his torso against the edge of his desk are used to determine the transformation such that the edge of the user's desk maps to the edge of the conference room table where the avatar is placed.

At step S72, data is stored, for example in memory 106, defining the relationship between each of the avatars which will be displayed to the user (that is, the avatars of the other participants) and the horizontal position on the display screen of monitor 34 at which the avatar will be displayed. As described above with respect to step S68, the avatars are positioned in the conference room model such that the position at which each avatar will appear across the display screen in a horizontal direction when an image is rendered is fixed. Accordingly, in this embodiment, data defining these fixed positions for each different number of participants is pre-stored in memory 106, and, at step S72, the data defining the fixed positions for the correct number of participants is selected and each of the fixed positions is assigned a participant number (received from the

conference coordinator at step S40) defining the participant displayed at that position. More particularly, as will now be described with reference to Figure 24, data defining a piece-wise linear function between the fixed positions of the avatars is stored and the participant numbers are associated with this data at step S72.

Referring to Figure 24, data for the display of six avatars is shown (corresponding to the examples described previously with respect to Figure 23D and Figure 23E). The vertical axis in Figure 24 shows horizontal screen position, and values on this axis range from -0.5 (corresponding to a position on the left hand edge of the screen) to +0.5 (corresponding to a position on the right hand edge of the screen). The horizontal axis has six equally spaced divisions 400, 402, 404, 406, 408 and 410, each of which corresponds to a participant. Accordingly, the value of the function at each of these positions on the horizontal axis is -0.46, -0.34, -0.12, +0.12, +0.34 and +0.46 respectively (as shown by the dots in Figure 24) since these are the horizontal screen positions at which the avatars for six participants will be displayed. Data is also stored defining a piece-wise linear function between each of these values. At step S72, each of the six positions on the horizontal

axis is assigned a participant number corresponding to the participant whose avatar will be displayed at the associated horizontal screen position. Referring to the seating plane shown in Figure 6, in this example, position 400 is allocated participant number 2, position 402 is allocated participant number 3, position 404 is allocated participant number 4, position 406 is allocated participant number 5, position 408 is allocated participant number 6 and position 410 is allocated participant number 7. It should be noted that the participant numbers for each of these positions will be different for each user station. By way of example, at the user station for participant 2, the participant numbers allocated to positions 400, 402, 404, 406, 408 and 410 will be 3, 4, 5, 6, 7 and 1 respectively.

As a result of allocating the participant numbers, the piece-wise linear function therefore defines, for each horizontal screen position a so-called "view parameter" V for the user which defines which participant in the conference room the user is looking at when he is looking at a particular position on the display screen of monitor 34. As will be explained below, during the video conference, processing is carried out to determine the horizontal position on the display screen which the user

is looking, and this is used to read the "view parameter" V for the user, which is then transmitted to the other participants to control the user's avatar.

At step S74, a three-dimensional computer model is set-up comprising a circular conference room table and name labels showing the name of each of the participants, and is stored in the meeting archive database 126 for archive purposes. More particularly, a three-dimensional computer model of a conference room table is prestored and, in step S74, name labels showing the name of each of the participants are placed around the conference room table at equal spacing in accordance with the seating plan received from the conference coordinator at step S40.

At step S76, a respective transformation is defined for each participant which transforms the avatar for the participant from the local coordinate system within the meeting archive database 126 in which it was stored at step S40 into the three-dimensional computer model created at step S74 so that the avatar appears at the correct position at the conference room table. As in step S70, in step S76, the three-dimensional positions of the body marker 72 previously received from each participant (as transmitted at step S282 in Figure 17)

when the participant was sitting with his torso against the edge of his desk are used to determine the transformation such that the edge of the user's desk maps to the edge of the conference room table where the avatar is placed.

Referring again to Figure 7, at step S78, when all of the preceding steps in Figure 7 have been completed, a "ready" signal is transmitted to the conference coordinator indicating that the user station has been calibrated and is now ready to start the video conference.

Referring again to Figure 4, at step S8, the video conference itself is carried out.

Figure 25 shows the processing operations which are performed to carry out the video conference.

Referring to Figure 25, the processes at steps S370, S372, S374-1 to S374-6, S376, S378 and S380 are carried out simultaneously and repeatedly.

At step S370, frames of image data are recorded by cameras 26 and 28 as the user participates in the video conference, that is as the user views the images of the

avatars of the other participants on monitor 34, listens to the sound data from the other participants and speaks into microphone 52. Synchronous frames of image data (that is, one frame from each camera which were recorded at the same time) are processed by image data processor 104 at video frame rate to generate in real-time data defining the three-dimensional coordinates of the body markers 70, 72, the view parameter V defining where the user was looking in the conference room when the images were recorded, and pixel data for the face of the user. This data is then transmitted to all of the other participants. Step S370 is repeated for subsequent pairs of frames of image data until the video conference ends.

Figure 26 shows the processing operations performed at step S370 for a given pair of synchronised frames of image data.

Referring to Figure 26, at step S390, synchronous frames of image data are processed to calculate the three-dimensional coordinates of the headset LEDs 56, 58, 60, 62, 64 and body markers 70, 72 which are visible in both of the images. This step is carried out in the same way as step S234 in Figure 14, and described above with respect to Figures 15 and 16, except that the processing

is performed for the body markers 70, 72 in addition to the headset LEDs. Accordingly, this processing will not be described again here.

At step S392, the plane of the user's head is determined by finding the plane which passes through the three-dimensional positions of the headset LEDs calculated at step S390 and adjusting this plane by the headset offset angle θ previously determined at step S64 (Figure 7).

At step S394, a line is projected from the plane of the user's head in a direction perpendicular to this plane, and the intersection of the projected line with the display screen of monitor 34 is calculated. This is illustrated in Figures 27A, 27B and 27C.

Referring to Figure 27A, in this embodiment, the mid-point 220 of the line between the three-dimensional coordinates of the headset LEDs 58 and 62 is determined and a line 218 is projected from the calculated mid-point 220 perpendicular to the plane 224 of the user's head (which was calculated at step S392 by determining the plane 228 of the headset LEDs and adjusting this by the headset offset angle θ). As described above with respect to step S50 (Figure 7), the headset LEDs 58 and 62 are aligned with the user's eyes so that, in this

embodiment, the projected line 218 is not only perpendicular to the plane 224 of the user's head, but also passes through a point on this plane representative of the position of the user's eyes.

Referring to Figure 27B, the projected line 218 intersects the plane of the display screen of monitor 34 at a point 240. In step S394, the horizontal distance "h" shown in Figure 27C of the point 240 from the centre of the display screen (that is, the distance between the vertical line in the plane of the display screen on which point 240 lies and the vertical line in the plane of the display screen on which the centre point of the display lies) is calculated using the three-dimensional coordinates of the display screen previously determined at step S66 (Figure 7) during calibration.

Referring again to Figure 26, at step S396, the view parameter V defining where the user was looking when the frames of image data being processed were recorded is determined. More particularly, the ratio of the distance "h" calculated at step S394 to the width "W" of the display screen stored at step S48 (Figure 7) is calculated and the resulting value is used to read a value for the view parameter V from the data stored at step S72 during calibration. By way of example, if the

distance "h" is calculated to be 2.76 inches and the width "W" of the display screen is 12 inches (corresponding to a 15 inch monitor), then a ratio of 0.23 would be calculated and, referring to Figure 24, this would cause a view parameter "V" of 5.5 to be generated. As can be seen from the example shown in Figures 27B and 27C, the projected ray 218 indicates that the user 44 is looking between participants 5 and 6, and hence a view parameter of 5.5 would define this position.

Referring again to Figure 26, at step S398, the direction of the imaging plane of each of the cameras 26 and 28 (that is, the plane in which the CCD of the camera lies) is compared with the direction of the plane of the user's head calculated at step S392 to determine which camera has an imaging plane most parallel to the plane of the user's head. Referring again to Figure 27B, for the example illustrated, it will be seen that the imaging plane 250 for camera 28 is more parallel to the plane 224 of the user's head than the imaging plane 252 of camera 26. Accordingly, in the example illustrated in Figure 27B, camera 28 would be selected at step S398.

At step S400, the frame of image data from the camera selected at step S398 is processed to extract the pixel data representing the user's face in the image. In this

embodiment, this step is performed using the three-dimensional positions of the headset LEDs 56 and 64 calculated at step S390, the size and ratio of the user's head determined at step S60 (Figure 7) and the distance "a" between each LED 56, 64 and the inner surface of the corresponding earpiece 48, 50 (which, as noted above, is pre-stored in PC 24). More particularly, using the three-dimensional positions of the headset LEDs 56 and 64, and the distance "a", the points representing the extents of the width of the user's head in three dimensions are determined. These extent points are then projected back into the image plane of the camera selected at step S398 using the camera transformation determined at step S62 (Figure 7). The projected points represent the extents of the width of the user's head in the image, and, using the value of this width and the ratio of the user's head length, the extents of the user's head length in the image are determined. Pixels representing the image between the extents of the width of the user's head and the extents of the length of the user's head are then extracted. In this way, image data is not extracted which shows the headset 30 which the user is wearing.

At step S401, the three-dimensional coordinates of the body markers 70, 72 calculated at step S390 are

transformed into the standardised coordinate system previously defined at step S66 in Figure 7.

At step S402, MPEG 4 encoder 108 encodes the face pixel data extracted at step S400, the 3D coordinates of the body markers 70, 72 generated at step S401 and the view parameter determined at step S396 in accordance with the MPEG 4 standard. More particularly, the face pixel data and the 3D coordinates are encoded as a Movie Texture and Body Animation Parameter (BAP) set and, since the MPEG 4 standard does not directly provide for the encoding of a view parameter, this is encoded in a general user data field. The encoded MPEG 4 data is then transmitted to the user stations of each of the other participants via input/output interface 110 and the Internet 20.

Referring again to Figure 25, at step S372, sound produced by user 44 is recorded with microphone 52 and encoded by MPEG 4 encoder 108 in accordance with the MPEG 4 standard. In addition, the text corresponding to the user's speech generated by voice recognition processor 107 is also encoded as a text stream in accordance with the MPEG 4 standard. The encoded sound and text is then transmitted to the other participants by input/output interface 110 and the Internet 20.

At steps S374-1 to S374-6, MPEG decoder 112, model processor 116 and central controller 100 perform processing to change the avatar models stored in avatar and 3D conference model store 114 in dependence upon the MPEG 4 encoded data received from the other participants. More particularly, in step S374-1 processing is performed to change the avatar of the first external participant using the data received from that participant, in step S374-2 the avatar of the second external participant is changed using data received from the second external participant etc. Steps S374-1 to S374-6 are performed simultaneously, in parallel.

Figure 28 shows the processing operations performed in each of steps S374-1 to S374-6.

Referring to Figure 28, at step S420, MPEG 4 decoder 112 awaits further data from the participant whose avatar is to be updated. When data is received, it is decoded by the MPEG 4 decoder, and the decoded data is then passed to model processor 116 at step S422, where it is read to control subsequent processing by model processor 116 and central controller 100.

At step S424, the position of the avatar body and arms are changed in the three-dimensional coordinate system

in which it is stored in avatar and 3D conference model store 114 so that the body and arms of the avatar fit the received three-dimensional coordinates of the body markers 70, 72 of the actual participant. In this way, the pose of the avatar is made to correspond to the real-life pose of the actual participant which the avatar represents.

At step S426, the face pixel data in the bitstream received from the participant is texture mapped onto the face of the avatar model in three dimensions.

At step S428, the avatar is transformed from the local coordinate system in which it is stored into the three-dimensional model of the conference room using the transformation previously defined at step S70 (Figure 7).

At step S430, the head of the transformed avatar in the three-dimensional conference room model is changed in dependence upon the view parameter, V, of the participant defined in the received bitstream. More particularly, the head of the avatar is moved in three dimensions so that the avatar is looking at the position defined by the view parameter. For example, if the view parameter, V, is 5, then the avatar's head is moved so that the avatar is looking at the position in the three-dimensional

conference room at which participant 5 is seated. Similarly, if, for example, the view parameter is 5.5, then the avatar's head is rotated so that the avatar is looking mid-way between the positions at which the fifth and sixth participants sit in the three-dimensional conference room.

Figures 29A, 29B and 29C illustrate how the position of the avatar's head is changed in the conference room model in dependence upon changes of the participant's head in real-life.

Referring to Figure 29A, an example is shown in which participant 1 in real-life is initially looking at participant 2 (or more particularly, the avatar of participant 2) on the display screen of his monitor, and then rotates his head through an angle β1 to look at participant 7 on the display screen. In real-life, the angle of rotation β1 would be approximately 20°-30° for typical screen sizes and seating positions from the screen.

Figure 29B represents the images seen by participant 3 of the video conference. When the head of participant 1 in real-life is looking at participant 2, then the head of the avatar 300 of participant 1 is positioned so that

it, too, is looking at the avatar of participant 2 in the three-dimensional model of the conference room stored at the user station of participant 3. As the first participant rotates his head in real-life to look at participant 7, the head of the avatar 300 undergoes a corresponding rotation to look at the avatar of participant 7 in the three-dimensional conference room model. However, the angle $\beta 2$ through which the head of avatar 300 moves is not the same as angle $\beta 1$ through which the head of the first participant moves in real-life. In fact, in this example, the angle $\beta 2$ is much larger than the angle $\beta 1$ due to the relative positions of the avatars in the conference room model. Consequently, the motion of the heads of the avatars does not take place in the same coordinate system as that of the motion of the heads of the actual participants in real-life.

The change in angle of the head of avatar 300 will be different for each user station since the arrangement of the avatars in the three-dimensional conference room model is different at each user station. Figure 29C illustrates how the head of avatar 300 moves in the image displayed at the user station of participant 2 as participant 1 moves his head in real-life through the angle $\beta 1$ to look from participant 2 to participant 7.

Referring to Figure 29C, since participant 1 is originally looking at participant 2, the head of avatar 300 is originally directed towards the viewing position from which the image is rendered for display to participant 2. As participant 1 rotates his head through angle β1 in real-life, the head of avatar 300 is rotated through angle β3 so that the head is looking at the avatar of participant 7 in the three-dimensional model of the video conference room stored at the user station of participant 2. The angle β3 is different to both β1 and β2.

Referring again to Figure 25, at step S376, image renderer 118 and central controller 100 generate and display a frame of image data on monitor 34 showing the current status of the three-dimensional conference room model and the avatars therein. The processing performed at step S376 is repeated to display images at video rate, showing changes as the avatars are updated in response to changes of the participants in real-life.

Figure 30 shows the processing operations performed at step S376.

Referring to Figure 30, at step S450, an image of the three-dimensional conference room model stored in avatar

and 3D conference model store 114 is rendered in a conventional manner to generate pixel data, which is stored in frame buffer 120.

At step S452, the current view parameter V determined at step S370 in Figure 25 (which occurs in parallel) is read. As noted above, this view parameter defines the position on the monitor at which the user is determined to be looking, relative to the avatars displayed.

At step S454, the image data generated and stored at step S450 is amended with data for a marker to show the position at which the user is determined to be looking in accordance with the view parameter read at step S452.

At step S456, the pixel data now stored in frame buffer 120 is output to monitor 34 to display an image on the display screen.

Figure 31 illustrates the display of markers in accordance with the users current view parameter V.

Referring to Figure 31, if for example it is determined at step S452 that the user's current view parameter is 5, then at step S454, image data for arrow 310 is added so that, when the image is displayed at step S456, the

user sees arrow 310 indicating that he is determined to be looking at participant 5 and that this is the information which will be transmitted to all of the other participants. Accordingly, if the displayed marker does not accurately indicate the user's intended viewing direction, the user can change the position of his head whilst watching the position of the marker change until the correct viewing direction is determined and transmitted to the other users.

By way of further example, if the user's view parameter is 6.5, then arrow 320 would be displayed (instead of arrow 310) indicating a position mid-way between the avatars of participants 6 and 7.

Referring again to Figure 25, at step S378, MPEG 4 decoder 112, central controller 100 and sound generator 122 perform processing to generate sound for the user's headset 30.

Figure 32 shows the processing operations performed at step S378.

Referring to Figure 32, at step S468 the input MPEG 4 bitstreams received from each participant are decoded by MPEG 4 decoder 112 to give a sound stream for each

participant.

At step S470, the current head position and orientation for each avatar in the coordinate system of the three-dimensional computer model of the conference room are read, thereby determining a sound direction for the sound for each of the avatars.

At step S472, the current head position and orientation of the user (to whom the sound will be output) is read (this having being already determined at step S370 in Figure 25), thereby defining the direction for which the output sound is to be generated.

At step S474, the input sound streams decoded at step S468, the direction of each sound stream determined at step S470 and the output direction for which sound is to be generated determined at step S472 are input to the sound generator 122, where processing is carried out to generate left and right output signals for the user's headset 30. In this embodiment, the processing in sound generator 122 is performed in a conventional manner, for example such as that described in "The Science of Virtual Reality and Virtual Environments" by R.S. Kalawsky, Addison-Wesley Publishing Company, ISBN 0-201-63171-7, pages 184-187.

In the processing described above, at step S472, the user's current head position and orientation are used to determine an output direction which is subsequently used in the processing of the sound streams at step S474. In this way, the sound which is output to the headset 30 of the user changes in dependence upon the user's head position and orientation, even though the images which are displayed to the user on monitor 34 do not change as his head position and orientation change (other than the displayed marker indicating where the user is looking).

Referring again to Figure 25, at step S380, archive processor 124 and central controller 100 process the decoded MPEG 4 bitstreams received from each of the other participants and also the user's own decoded MPEG 4 bitstream to update the data stored in the meeting archive database 126.

Figure 33 shows the processing operations performed at step S380.

Referring to Figure 33, at step S400, archive processor 124 and central controller 100 use the decoded view parameter and the 3D coordinates of the body markers from each decoded MPEG 4 bitstream to update the three-dimensional computer model stored in the meeting archive

database 126.

Figure 34 shows the processing operations performed at step S400.

Referring to Figure 34, at step S430, the position of the body and arms of each avatar are changed in the respective three-dimensional coordinate systems in which they were stored in meeting archive database 126 so that the body and arms of the avatars fit the received three-dimensional coordinates of the body markers 70, 72 of the actual participants. In this way, the pose of each avatar is made to agree with the real-life pose of the corresponding actual participant.

At step S432, each avatar is transformed from its own local coordinate system into the three-dimensional conference model using the transformations previously calculated at step S76 (Figure 7). This results in a three-dimensional computer model in which the avatars of the participants are spaced evenly around a circular conference room table in the seating order defined in the data received from the conference coordinator at step S40 (Figure 7). An example of such a model is shown in Figure 35.

At step S434, the head of each transformed avatar in the three-dimensional conference model is changed in dependence upon the associated view parameter, V, defined in the received bitstream. More particularly, the head of each avatar is moved in three-dimensions so that the avatar is looking at the position defined by the view parameter.

Referring again to Figure 33, at step S402, data defining the updated meeting model produced at step S400, the face pixel data for each participant from the decoded MPEG 4 bitstreams and the sound data of each participant from the decoded MPEG 4 bitstreams is encoded as MPEG 4 data and stored in the meeting archive database 126.

Figures 36A and 36B schematically illustrate the storage of data in meeting archive database 126. The storage structure shown in Figures 36A and 36B is notional and is provided for illustration purposes only to assist understanding; it does not necessarily represent the exact way in which data is stored in meeting archive database 126.

Referring to Figures 36A and 36B, meeting archive database 126 stores data 480 defining the relationship between participant number and participant name, and data

490 comprising an archive record of the video conference. The data 490 includes timing information represented by the horizontal axis 500, on which each unit represents a predetermined amount of time, for example one frame of MPEG 4 data. The updated meeting model, face pixel data and sound data encoded at step S402 is stored as MPEG 4 data 502 in the database, together with timing information (this timing information being schematically represented in Figure 36B by the position of the MPEG 4 data 502 along the horizontal axis 500).

Referring again to Figure 33, at step S404, the view parameter, V, of each participant is quantised. More particularly, in this embodiment, the quantisation is carried out such that, if a viewing parameter is within 0.25 of a participant number, then the viewing parameter is quantised to be equal to the viewing participant number, otherwise it is determined that the corresponding participant is not looking at any of the other participants in particular, and accordingly the viewing parameter is set to "0". That is, quantisation is performed using the following equations:

$$If\ i-0.25 \leq V \leq i+0.25\ then\ set\ V=i \qquad (11)$$

$$Otherwise\ set\ V=0 \qquad (12)$$

where "i" is an integer participant number (i=1,2, 3, ...).

At step S406, the decoded text stream of each participant is checked to determine whether any text is present, indicating that the participant is speaking, and, at step S408, the quantised viewing parameter produced at step S404 is stored for each participant that is speaking for subsequent analysis, for example in cache memory.

At step S410, the decoded text data for each participant is stored in the meeting archive database 126 (indicated at 504 in Figure 36B). As will be appreciated, however, there will not be any text data for storage from participants who are not speaking. In the example shown in Figure 36B, text is stored for the first five time slots for participant 1 (indicated at 506), for the seventh to twelfth time slots for participant 4 (indicated at 508), and for the eleventh to twentieth time slots for participant 5 (indicated at 510). No text is stored for participants 2, 3, 6 and 7 since none of these participants spoke during the time slots in the example shown in Figure 36B.

At step S412, the quantised viewing parameters calculated at step S404 are stored in the meeting archive database

126. Referring to Figure 36B, the stored quantised viewing parameters are indicated at 512. For each participant, a quantised viewing parameter is stored for each time slot. Thus, referring to the quantised viewing parameters of participant 1 by way of example, the parameters indicate that participant 1 was looking at participant 3 during the first two time slots, participant 4 for the next three time slots, and at none of the other participants for the following two time slots etc.

At step S414, central controller 100 and archive processor 124 determine whether one or more of the participants has stopped speaking. In this embodiment, this check is performed by examining the text data 504 to determine whether text data for a given participant was present for the previous time slot, but is not present for the current time slot. If this condition is satisfied for any participant (that is, a participant has stopped speaking), then, at step S416, the quantised viewing parameter values previously stored at step S408 are processed for each participant which has stopped speaking to generate data defining a viewing histogram. More particularly, the quantised viewing parameters for the period in which the participant was speaking are processed to generate data defining the percentage of

time during that period that the speaking participant was looking at each of the other participants.

Figures 37A, 37B and 37C show the viewing histograms corresponding to the periods of text 506, 508 and 510 in Figure 36B, respectively. Referring to Figure 36B and Figure 37A, during the period 506 when participant 1 was speaking, he was looking at participant 3 for two of the five time slots (that is, 40% of the total length of the period for which he was talking), which is indicated at 520 in Figure 37A, and at participant 4 for three of the five time slots (that is, 60% of the time), which is indicated at 524 in Figure 37A.

Similarly, during period 508, participant 4 was looking at participant 1 for approximately 83% of the time, which is indicated at 526 in Figure 37B, and at user 5 for approximately 17% of the time, which is indicated at 528 in Figure 37B.

During period 510, participant 5 was looking at user 1 for 70% of the time, indicated at 530 in Figure 37C, and at participant 4 for 30% of the time, indicated at 532 in Figure 37C.

Referring again to Figure 33, at step S418, the viewing

histograms generated at step S416 are stored in the meeting archive database 126 linked to the associated period of text for which they were generated. Referring to Figure 36B, the stored viewing histograms are indicated at 540, with the data defining the histogram for text period 506 indicated at 552, the data defining the histogram for text period 508 indicated at 554, and the data defining the histogram for text period 510 indicated at 556.

On the other hand, if it is determined at step S414 in Figure 33 that, for the current time period, one of the participants has not stopped speaking, then steps S416 and S418 are omitted.

Referring again to Figure 4, at step S10, the meeting archive database 126 is interrogated to retrieve data relating to the meeting.

Figure 38 shows the processing operations performed at step S10.

Referring to Figure 38, at step S450, central controller 100 causes a message to be displayed on display device 34 prompting the user to enter information defining the search of meeting archive database 126 which is required.

More particularly, in this embodiment, central controller 100 causes the display shown in Figure 39A to appear on display device 34.

Referring to Figure 39A, the user is requested to enter information defining the part or parts of the meeting which he wishes to find in the meeting archive database 126. More particularly, in this embodiment, the user is requested to enter information 600 defining a participant who was talking, information 610 comprising one or more key words which were said by the participant identified in information 600, and information 620 defining the participant to whom the participant identified in information 600 was talking. In addition, the user is able to enter time information defining a portion or portions of the meeting for which the search is to be carried out. More particularly, the user can enter information 630 defining a time in the meeting beyond which the search should be discontinued (that is, the period of the meeting before the specified time should be searched), information 640 defining a time in the meeting after which the search should be carried out, and information 650 and 660 defining a start time and end time respectively between which the search is to be carried out. In this embodiment, information 630, 640, 650 and 660 may be entered either by specifying a time

in absolute terms, for example in minutes, or in relative terms by entering a decimal value which indicates a proportion of the total meeting time. For example, entering the value 0.25 as information 630 would restrict the search to the first quarter of the meeting.

In this embodiment, the user is not required to enter all of the information 600, 610 and 620 for one search, and instead may omit one or two pieces of this information. If the user enters all of the information 600, 610 and 620, then the search will be carried out to identify each part of the meeting in which the participant identified in information 600 was talking to the participant identified in information 620 and spoke the key words defined in information 610. On the other hand, if information 610 is omitted, then a search will be carried out to identify each part of the meeting in which the participant defined in information 600 was talking to the participant defined in information 620 irrespective of what was said. If information 610 and 620 is omitted, then a search is carried out to identify each part of the meeting in which the participant defined in information 600 was talking, irrespective of what was said and to whom. If information 600 is omitted, then a search is carried out to identify each part of the meeting in which any of the participants spoke the key words defined in

information 610 to the participant defined in information 620. If information 600 and 610 is omitted, then a search is carried out to identify each part of the meeting in which any of the participants spoke to the participant defined in information 620. If information 620 is omitted, then a search is carried out to identify each part of the meeting in which the participant defined in information 600 spoke the key words defined in information 610, irrespective of to whom the key word was spoken. Similarly, if information 600 and 620 is omitted, then a search is carried out to identify each part of the meeting in which the key words identified in information 610 were spoken, irrespective of who said the key words and to whom.

In addition, the user may enter all of the time information 630, 640, 650 and 660 or may omit one or more pieces of this information.

Once the user has entered all of the required information to define the search, he begins the search by clicking on area 670 using a user input device such as the mouse 38.

Referring again to Figure 38, at step S452, the search information entered by the user is read by central

controller 100 and the instructed search is carried out. More particularly, in this embodiment, central controller 100 converts any participant names entered in information 600 or 620 to participant numbers using the table 480 (Figure 36A), and considers the text information 504 for the participant defined in information 600 (or all participants if information 600 is not entered). If information 620 has been entered by the user, then, for each period of text, central controller 100 checks the data defining the corresponding viewing histogram to determine whether the percentage of viewing time in the histogram for the participant defined in information 620 is equal to or above a threshold which, in this embodiment, is 25%. In this way, periods of speech (text) are considered to satisfy the criteria that a participant defined in information 600 was talking to the participant defined in information 620 even if the speaking participant looked at other participants while speaking, provided that the speaking participant looked at the participant defined in information 620 for at least 25% of the time of the speech. Thus, the period of speech in which the value of the viewing histogram is equal to or above 25% for two or more participants would be identified if any of these participants were specified in information 620. If the information 610 has been input by the user, then central controller 100 and text

searcher 128 search each portion of text previously identified on the basis of information 600 and 620 (or all portions of text if information 600 and 620 was not entered) to identify each portion containing the key word(s) identified in information 610. If any time information has been entered by the user, then the searches described above are restricted to the meeting times defined by those limits.

At step S454, central controller 100 displays a list of relevant speeches identified during the search to the user on display device 34. More particularly, central controller 100 displays information such as that shown in Figure 39B to the user. Referring to Figure 39B, a list is produced of each speech which satisfies the search parameters, and information is displayed defining the start time for the speech both in absolute terms and as a proportion of the full meeting time. The user is then able to select one of the speeches for playback by clicking on the required speech in the list using the mouse 38.

At step S456, central controller 100 reads the selection made by the user at step S454, and plays back the stored MPEG 4 data 502 for the relevant part of the meeting from meeting archive database 126. More particularly, central

controller 100 and MPEG 4 decoder 112 decode the MPEG 4 data 502 and image renderer 118, frame buffer 120 and sound generator 122 are used to output the image data and sound. The image data comprises images rendered from the three-dimensional conference model which comprises the conference room table with the participant avatars equally spaced around it. This model may be viewed from any viewing position and direction by the user, by inputting instructions using the user input devices 36, 38, and the viewing position and/or direction can be changed during viewing.

A number of modifications are possible to the embodiment of the invention described above.

For example, in the embodiment described above, the cameras 26 and 28 at each user station record images of a single user at the user station and processing is performed to determine transmission data for the single user. However, the cameras 26 and 28 may be used to record images of more than one user at each user station and processing may be carried out to generate the face pixel data, the three-dimensional coordinates of the body markers and the view parameter for each of the users at the user station, and to transmit this data to the other participants to facilitate the animation of an avatar

corresponding to each one of the users.

In the embodiment above at steps S42 and S44 (Figure 7), camera parameters are input by the user. However, each of the cameras 26, 28 may be arranged to store these parameters and to pass it to PC 32 when the camera is connected to the PC.

In the embodiment above, LEDs 56, 58, 60, 62 and 64 are provided on headset 30. However, other forms of lights or identifiable markers may be provided instead.

In the embodiment described above, the headset LEDs 56, 58, 60, 62, 64 are continuously illuminated and have different colours to enable them to be identified in an image. Instead of having different colours, the LEDs could be arranged to flash at different rates to enable them to be distinguished by comparison of images over a plurality of frames, or the LEDs may have different colours and be arranged to flash at different rates.

In the embodiment above, the coloured body markers 70, 72 may be replaced by LEDs. Also, instead of using coloured markers or LEDs, the position of the user's body may be determined using sensors manufactured by Polhemus Inc., Vermont, USA, or other such sensors.

In the embodiment above, in the processing performed at step S370 (Figure 25) data for the whole of each image is processed at step S390 (Figure 26) to determine the position of each LED and each coloured body marker in the image. However, the position of each LED and each body marker may be tracked through successive frame of image data using conventional tracking techniques, such as Kalman filtering techniques, for example as described in "Affine Analysis of Image Sequences" by L.S. Shapiro, Cambridge University Press, 1995, ISBN 0-521-55063-7, pages 24-34.

In the embodiment above, at step S72 (Figure 7), data is stored defining the relationship between horizontal screen position and the view parameter V. Further, at step S396 (Figure 26), this stored data is used to calculate the view parameter to be transmitted to the other participants in dependence upon the horizontal distance between the point on the display screen at which the user is looking and the centre of the display screen. This method of determining the view parameter V is accurate when the viewing position from which the 3D model of the conference room and avatars is rendered is such that the participants are displayed to the user with their heads at substantially the same vertical height on the screen. However, errors can occur when the viewing

position is such that the heads of the participants are at different heights on the display screen. To address this, it is possible to store data at step S72 defining the relationship between the view parameter V and the distance of each avatar around the arc 164 (from any fixed point), and at step S396 to calculate the point on arc 164 which is nearest to the point on the screen at which the user is looking and use the calculated point on arc 164 to read the view parameter V which is to be transmitted to the other participants from the stored data. Further, although in the embodiment above the viewing position from which the 3D conference room model and avatars are rendered is fixed, it is possible to allow the user to vary this position. The view parameter V would then be calculated most accurately using the positions of the avatars around arc 164 as described above.

In the embodiment above, in the processing performed at step S370 (Figure 25), the user's view parameter is determined in dependence upon the orientation of the user's head. In addition, or instead, the orientation of the user's eyes may be used.

In the embodiment above, the sound from the user's own microphone 52 is fed to the user's headphones 48, 50.

However, the user may be able to hear his own voice even when wearing the headphones, in which case such processing is unnecessary.

In the processing performed at step S62 (Figure 7) in the embodiment above, both a perspective camera transformation and an affine transformation are calculated and tested (steps S130 and S132 in Figure 9). However, it is possible to calculate and test just an affine transformation and, if the test reveals acceptable errors, to use the affine transformation during the video conference, or, if the test reveals unacceptable errors, to calculate and use a perspective transformation.

In the embodiment above, the names of the participants displayed on the name plates are based on the information provided by each participant to the conference coordinator at step S20 (Figure 5). However, the names may alternatively be based on other information, such as the log-on information of each participant at a user station, the telephone number of each user station, or information provided in the data defining the avatar of each participant.

In the embodiment above, at step S68 (Figure 7), the positions of the avatars around the conference room table

in the avatar and 3D conference model store 114 are set using the values given in Table 1. However, other positions may be used. For example, the avatars may be arranged so that their horizontal positions on the display screen are given by the following equation:

$$W_n = 0.46W \cos\left(\frac{180i}{N-1}\right) \quad \ldots.(13)$$

where: N is the number of avatars displayed on the screen

W_n is the position of the nth avatar (n = 1 ...N)

i = n-1

W is the screen width

In the embodiment above, at step S400 (Figure 26), the face pixel data is extracted following processing to determine the extents of the user's head such that the extracted pixel data will not contain pixels showing the headset 30. Instead, the pixel data may be extracted from an image by simply extracting all data bounded by the positions of the LEDs 56, 60 and 64 and using the user's head ratio to determine the data to extract in the direction of the length of the user's face. Conventional image data interpolation techniques could then be used to amend the pixel data to remove the headset 30.

In the embodiment above, a view parameter V is calculated to define the position of the head of an avatar. In this way, movements of the user's head in real-life are appropriately scaled to give the correct movement of the avatar's head in the three-dimensional conference room models in the avatar and 3D conference model store 114 and the meeting archive database 126 at the user stations of the other participants. In addition, it is also possible to perform corresponding processing for user gestures, such as when the user points, nods his head, etc. at a particular participant (avatar) on his display screen.

In the embodiment above, two cameras 26 and 28 are used at each user station to record frames of image data of the user 44. The use of two cameras enables three-dimensional position information to be obtained for the headset LEDs and body markers. However, instead, a single camera could be used together with a range finder to provide depth information. Further, a single calibrated camera could be used on its own, with depth information obtained using a standard technique, for example as described in "Computer and Robot vision, Volume 2" by R.M. Haralick and L.G. Shapiro, Addison-Wesley Publishing Company, 1993, ISBN 0-201-56943-4, pages 85-91.

Instead of using LEDs or coloured markers to determine the position of the user's head, arms and torso, conventional feature matching techniques could be used to match natural features of the user in each of the images in a pair of synchronised images. Examples of conventional techniques are given in "Fast visual tracking by temporal consensus" by A.H. Gee and R. Cipolla in Image and Vision Computing, 14(2): 105-114, 1996, in which nostrils and eyes are tracked and "Learning and Recognising Human Dynamics in Video Sequences" by C. Bregler, Proceedings IEEE Conference on Computer Vision and Pattern Recognition, June 1997, pages 568-574, in which blobs of motion and colour similarity corresponding to arms, legs and torso are tracked.

In the embodiment above, the view parameter, V, for a given participant defines at which other participant the participant is looking. However, the view parameter may also be used to define at which object the participant is looking, for example a display board or the participant's notes etc. Thus, in Figure 39A, information 620 would be used to specify at whom or at what the participant was looking when he was talking.

In the embodiment above, at step S404, the view parameter, V, of each participant is quantised using

equations 11 and 12. The view parameter may be quantised in different ways, however. For example, the view parameter may be quantised taking into account the position at which the participant is looking on his display screen and the distance between the avatars displayed on the display screen. More particularly, referring to Figure 40, the distance "d" between the point 240 at which it is determined that the participant is looking on the display screen and the centre of the head of the nearest avatar 194 is calculated, together with the distance "l" between the centre of the head of the nearest avatar 194 and the centre of the head of the avatar 196 which is the second nearest avatar to the point 240 at which the user is looking. Quantisation of the view parameter, V, is then calculated using the following equations:

$$If\ d < 0.3l\ then\ set\ V=i \qquad (14)$$

$$Otherwise\ set\ d=0 \qquad (15)$$

where i is the participant number of the nearest avatar.

In the embodiment above, the three-dimensional conference model stored in meeting archive database 126 is different to the three-dimensional model stored in avatar and 3D conference model store 114. However, the same computer model may be used both during the video conference and

for meeting archive purposes.

In the embodiment above, at step S452 (Figure 38), the viewing histogram of a particular portion of text is considered and it is determined that the participant was talking to a further participant if the percentage of gaze time for the further participant in the viewing histogram is equal to or above a predetermined threshold. Instead however, rather than using a threshold, the participant to whom the speaking participant was looking during the period of text may be defined to be the participant having the highest percentage gaze value in the viewing histogram (for example participant 4 in Figure 37A, participant 1 in Figure 37B and participant 1 in Figure 37C).

In the embodiment above, the text information produced by voice recognition processor 107 is encoded as an MPEG 4 text stream. However, the text data may be provided as binary asci code and encoded as MPEG 4 "user data".

In the embodiment above, the MPEG 4 data 502, the text data 504, the quantised view parameters 512 and the viewing histograms 540 are generated and stored in meeting archive database 126 in real-time as data is received from the other participants. However, instead,

the MPEG 4 data received from the other participants may be stored and the data 502, 504, 512 and 540 generated and stored in meeting archive database 126 in non-real-time.

In the embodiment above, the MPEG 4 data 502, the text data 504, the quantised view parameters 512 and the viewing histograms 540 are generated and stored in meeting archive database 126 before the database is interrogated to retrieve data for a defined part of the meeting. However, some, or all, of the data 502, 504, 512 and 540 may be generated in response to a search of the meeting archive database 126 being requested by the user, rather than being generated and stored prior to such a request. For example, in the embodiment above, at steps S400 and S402 (Figure 33), the 3D conference model in meeting archive database 126 is updated and the updated model is encoded as MPEG 4 data which is then stored as data 502 in meeting archive database 126. However, instead, the body animation parameters (defining the 3D coordinates of the body markers and the view parameter) may be stored for each participant together with the face pixel data and sound data and, during playback of a part of the meeting to the user, the stored body animation parameters may be used to change the positions of the avatars in the 3D model.

Similarly, although in the embodiment above the viewing histograms 540 are calculated and stored in real-time at steps S416 and S418, these histograms could be calculated in response to a search request being input by the user.

In the embodiment above, each user station 2, 4, 6, 8, 10, 12, 14 includes a voice recognition processor 107 which produces text data for output to the other user stations. In a further embodiment, voice data, but no text data may be sent from each user station to the other user stations, and the text data may be produced using a voice recognition processor in the receiving user station. In this embodiment, as part of the procedure shown in Figure 5, data defining the user's voice profile necessary to carry out voice recognition would be sent to the conference coordinator and then distributed to the other participants. Such user profile parameters are, for example, the parameters stored in the "user files" of Dragon Dictate. In further embodiments, audio data may be stored in meeting archive database 126 instead of the text data 504. The stored audio data would then either itself be searched for key words using voice recognition processing, or converted to text using voice recognition processing and the text searched using a conventional text searcher.

Instead of interconnecting the user stations 2, 4, 6, 8, 10, 12, 14 as in the embodiment described above, a client-server type of system may be used in which all communication between the user stations is directed through a server (which could, itself, be one of the user stations). In such an arrangement, the server could be arranged to store and maintain the meeting archive database 126 and to perform all searches thereon, rather than providing this facility in each user station.

The meeting archiving technique described above is applicable to all types of three-dimensional conferences and meetings. For example, it may be applied to a meeting which takes place in the same virtual space with each user wearing a virtual reality headset which provides information to define the view parameter, V.

The meeting archive technique described above can also be applied to data recorded at a single user station. More particularly, it can be used in a situation in which images are recorded of a plurality of users using one or more cameras and a three-dimensional computer model is amended in dependence upon the movements of the users. In such a case, the users could, for example be participating in a face-to-face meeting.

In the embodiment above, processing is performed by a computer using processing routines defined by programming instructions. However, some, or all, of the processing could be performed using hardware.

Different combinations of the above modifications are, of course, possible and other changes and modifications can be made without departing from the spirit and scope of the invention.

The contents of the applicant's co-pending applications filed concurrently herewith (attorney reference numbers: (GB9905197.1) (GB9905202.1) (GB9905158.3) (GB9905201.1) (GB9905166.4) (GB9905160.9) 2643401, 2643601, 2643701, 2643901, 2644001, 2644101, (GB9905199.7) (GB9905187.2) 2644201 and 2644601) are hereby incorporated by reference.

CLAIMS

1. Apparatus for archiving data defining models of three-dimensional objects, comprising:

receiving means for receiving data defining movements of the three-dimensional objects;

means for generating first archive data defining where at least one object is looking; and

a database for storing the first archive data such that it is associated with archive data for the three-dimensional objects.

2. Apparatus according to claim 1, wherein the means for generating first archive data is arranged to generate first archive data defining a person, animal or object at which the at least one object is looking.

3. Apparatus according to claim 1 or claim 2, wherein the database is arranged to store archive data for the three-dimensional objects comprising data defining the movements thereof.

4. Apparatus according to any preceding claim, wherein the database is arranged to store archive data for the three-dimensional objects comprising data defining the three-dimensional objects in different positions.

5. Apparatus according to any preceding claim, wherein the means for generating first archive data is arranged to generate respective first archive data for data received at different times by the receiving means, and the database is arranged to store the first archive data associated with the corresponding archive data for the three-dimensional objects and associated with second archive data comprising timing data.

6. Apparatus according to any preceding claim, further comprising means for generating third archive data defined text corresponding to words spoken by one of the three-dimensional objects, and wherein the database is arranged to store the third archive data such that it is associated with the archive data for the three-dimensional objects and related archive data.

7. Apparatus according to claim 6, wherein the means for generating third archive data comprises means for receiving audio data and means for processing the audio data to generate the text data therefrom.

8. Apparatus according to any preceding claim, wherein the receiving means is arranged to receive audio data, and wherein the database is arranged to store the audio data such that it is associated with the archive data for

the three-dimensional objects and related archive data.

9. Apparatus according to any preceding claim, wherein the database is arranged to store the archive data for the three-dimensional objects as MPEG 4 data.

10. Apparatus according to claim 9 when dependent upon claim 8, wherein the database is arranged to store the archive data for the three-dimensional objects and the audio data as MPEG 4 data.

11. Apparatus according to any preceding claim, further comprising means for generating fourth archive data defining, for a predetermined period, the proportion of time spend by a given three-dimensional object looking at each of other three-dimensional objects during the predetermined period, and wherein the database is arranged to store the fourth archive data such that it is associated with the archive data for the three-dimensional objects and related archive data.

12. Apparatus according to claim 11, wherein the predetermined period comprises a period during which the given three-dimensional object was talking.

13. Apparatus according to any preceding claim, further

comprising means for processing image data from at least one camera to generate data defining movements of the three-dimensional objects.

14. Apparatus according to any preceding claim, further comprising:

storage means for storing data defining three-dimensional models of the objects; and

means for changing the data in the storage means to give amended model data in dependence upon the data received by the receiving means defining movements of the objects;

and wherein the database is arranged to store the amended model data as the archive data for the three-dimensional objects.

15. Apparatus according to any preceding claim, further comprising search means for searching data stored in the database in accordance with search instructions to identify each part of the archive data for the three-dimensional objects which meets criteria defined in the search instructions.

16. Apparatus according to claim 15, wherein the search means is operable to search data stored in the database in accordance with search criteria relating to the first

archive data.

17. Apparatus according to claim 15 or claim 16, wherein the search means is operable to search data stored in the database in accordance with search criteria relating to any one or more of the first, second, third or fourth archive data.

18. Apparatus according to any preceding claim, further comprising means for rendering a model of the three-dimensional objects to generate image data showing the movements of the objects.

19. Apparatus for generating data for archiving data defining models of three-dimensional objects, comprising:

receiving means for receiving data defining movements of the three-dimensional objects; and

means for generating first archive data defining where at least one object is looking.

20. Apparatus according to claim 19, wherein the means for generating first archive data is arranged to generate first archive data defining a person, animal or object at which the at least one object is looking.

21. Apparatus according to claim 19 or claim 20, wherein

the means for generating first archive data is arranged to generate respective first archive data for data received at different times by the receiving means.

22. Apparatus according to any of claims 19 to 21, further comprising means for generating third archive data defined text corresponding to words spoken by one of the three-dimensional objects.

23. Apparatus according to claim 22, wherein the means for generating third archive data comprises means for receiving audio data and means for processing the audio data to generate the text data therefrom.

24. Apparatus according to any of claims 19 to 23, further comprising means for generating fourth archive data defining, for a predetermined period, the proportion of time spend by a given three-dimensional object looking at each of other three-dimensional objects during the predetermined period.

25. Apparatus according to claim 24, wherein the predetermined period comprises a period during which the given three-dimensional object was talking.

26. Apparatus according to any of claims 19 to 25,

further comprising means for processing image data from at least one camera to generate data defining movements of the three-dimensional models.

27. Apparatus according to any of claims 19 to 26, further comprising:

storage means for storing data defining three-dimensional models of the objects; and

means for changing the data in the storage means to give amended model data in dependence upon the data received by the receiving means defining movements of the objects.

28. Apparatus according to any of claims 19 to 27, further comprising means for rendering a model of the three-dimensional objects to generate image data showing the movements of the objects.

29. Apparatus for searching data stored in a database comprising data for three-dimensional objects and first archive data defining where the objects are looking, the apparatus comprising search means for searching the data in the database in accordance with search criteria relating to the first archive data.

30. Apparatus according to claim 29, wherein the data

for the three-dimensional objects comprises data defining movements thereof.

31. Apparatus according to claim 29, wherein the data for the three-dimensional objects comprises data defining the three-dimensional objects in different positions.

32. Apparatus according to any of claim 29 to 31, wherein the search means is further operable to search data in the database in accordance with search criteria relating to second archive data comprising data relating to the timing of the movements of the objects.

33. Apparatus according to any of claims 29 to 32, wherein the search means is further operable to search data in the database in accordance with search criteria relating to third archive data comprising text corresponding to words spoken by the objects.

34. Apparatus for archiving virtual meeting data, comprising:

storage means for storing data defining a three-dimensional model of the participants in the meeting;

means for receiving data defining movements of the participants;

means for generating archive data defining where

each person is looking; and

database means for storing the archive data linked to data defining the movements of the participants.

35. Apparatus for archiving virtual meeting data, comprising:

storage means for storing data defining a three-dimensional model of the participants in the meeting;

means for receiving data defining movements of the participants;

means for generating archive data defining where each person is looking; and

database means for storing the archive data linked to data defining the three-dimensional model amended in accordance with the data defining the movements.

36. Apparatus according to claim 34 or claim 35, wherein the receiving means is arranged to receive data defining how to amend data stored in the storage means to represent movements of three-dimensional models therein.

37. A method of archiving data defining models of three-dimensional objects, comprising:

receiving data defining movements of the three-dimensional objects;

generating first archive data defining where at

least one object is looking; and

storing the first archive data in a database such that it is associated with archive data for the three-dimensional objects.

38. Apparatus according to claim 37, wherein the step of generating first archive data comprises generating first archive data defining a person, animal or object at which the at least one object is looking.

39. A method according to claim 37 or claim 38, wherein archive data for the three-dimensional objects comprising data defining the movements thereof is stored in the database.

40. A method according to any of claims 37 to 39, wherein archive data for the three-dimensional objects comprising data defining the three-dimensional objects in different positions is stored in the database.

41. A method according to any of claims 37 to 40, wherein the step of generating first archive data comprises generating respective first archive data for data received at different times in the receiving step, and wherein the first archive data is stored in the database so as to be associated with the corresponding

archive data for the three-dimensional objects and so as to be associated with second archive data comprising timing data.

42. A method according to any of claims 33 to 41, further comprising the steps of generating third archive data defined text corresponding to words spoken by one of the three-dimensional objects, and storing the third archive data in the database such that it is associated with the archive data for the three-dimensional objects and related archive data.

43. A method according to claim 42, wherein the step of generating third archive data comprises receiving audio data and processing the audio data to generate the text data therefrom.

44. A method according to any of claims 37 to 43, wherein, in the receiving step, audio data is received, and wherein the audio data is stored in the database such that it is associated with the archive data for the three-dimensional objects and related archive data.

45. A method according to any of claims 37 to 44, wherein the archive data for the three-dimensional objects is stored in the database as MPEG 4 data.

46. A method according to claim 45 when dependent upon claim 44, wherein the archive data for the three-dimensional objects and the audio data is stored in the database as MPEG 4 data.

47. A method according to any of claims 37 to 46, further comprising the step of generating fourth archive data defining, for a predetermined period, the proportion of time spend by a given three-dimensional object looking at each of other three-dimensional objects during the predetermined period, and wherein the fourth archive data is stored in the database such that it is associated with the archive data for the three-dimensional objects and related archive data.

48. A method according to claim 47, wherein the predetermined period comprises a period during which the given three-dimensional object was talking.

49. A method according to any of claims 37 to 48, further comprising the step of processing image data from at least one camera to generate data defining movements of the three-dimensional objects.

50. A method according to any of claims 37 to 49, further comprising the step of changing data defining

three-dimensional models of the objects to give amended model data in dependence upon the data received in the receiving step defining movements of the objects, and wherein the amended model data is stored in the database as the archive data for the three-dimensional objects.

51. A method according to any of claims 37 to 50, further comprising a search step of searching data stored in the database in accordance with search instructions to identify each part of the archive data for the three-dimensional objects which meets criteria defined in the search instructions.

52. A method according to claim 51, wherein, in the search step, data stored in the database is searched in accordance with search criteria relating to the first archive data.

53. A method according to claim 51 or claim 52, wherein, in the search step, data stored in the database is searched in accordance with search criteria relating to any one or more of the first, second, third or fourth archive data.

54. A method according to any of claims 37 to 53, further comprising the step of rendering a model of the

three-dimensional objects to generate image data showing the movements of the objects.

55. A method according to any of claims 37 to 54, further comprising a step of generating a signal conveying the database with data therein.

56. A method according to claim 55, further comprising a step of recording the signal either directly or indirectly to generate a recording thereof.

57. A method of generating data for archiving data defining models of three-dimensional objects, comprising:

receiving data defining movements of the three-dimensional objects; and

generating first archive data defining where at least one object is looking.

58. A method according to claim 57, wherein the step of generating first archive data comprises generating first archive data defining a person, animal or object at which the at least one object is looking.

59. A method according to claim 57 or claim 58, wherein the step of generating first archive data comprises generating respective first archive data for data

received at different times in the receiving step.

60. A method according to any of claims 57 to 59, further comprising the step of generating third archive data defined text corresponding to words spoken by one of the three-dimensional objects.

61. A method according to claim 60, wherein the step of generating third archive data comprises receiving audio data and processing the audio data to generate the text data therefrom.

62. A method according to any of claims 57 to 61, further comprising the step of generating fourth archive data defining, for a predetermined period, the proportion of time spend by a given three-dimensional object looking at each of other three-dimensional objects during the predetermined period.

63. A method according to claim 62, wherein the predetermined period comprises a period during which the given three-dimensional object was talking.

64. A method according to any of claims 57 to 63, further comprising the step of processing image data from at least one camera to generate data defining movements

of the three-dimensional models.

65. A method according to any of claims 57 to 64, further comprising the step of changing data defining three-dimensional models of the objects to give amended model data in dependence upon the data received in the receiving step defining movements of the objects.

66. A method according to any of claims 57 to 65, further comprising the step of rendering a model of the three-dimensional objects to generate image data showing the movements of the objects.

67. A method according to any of claims 57 to 66, further comprising the step of generating a signal conveying the first archive data.

68. A method according to claim 67, further comprising the step of recording the signal either directly or indirectly to generate a recording thereof.

69. A method of searching data stored in a database comprising data for three-dimensional objects and first archive data defining where the objects are looking, wherein the data in the database is searched in accordance with search criteria relating to the first

archive data.

70. A method according to claim 69, wherein the data for the three-dimensional objects comprises data defining movements thereof.

71. A method according to claim 69, wherein the data for the three-dimensional objects comprises data defining the three-dimensional objects in different positions.

72. A method according to any of claim 69 to 71, wherein the database is searched in accordance with further search criteria relating to second archive data comprising data relating to the timing of the movements of the objects.

73. A method according to any of claims 69 to 72, wherein the database is searched in accordance with further search criteria relating to third archive data comprising text corresponding to words spoken by the objects.

74. A method of archiving virtual meeting data, comprising:

a storage step of storing data defining a three-dimensional model of the participants in the meeting;

receiving data defining movements of the participants;

generating archive data defining where each person is looking; and

storing the archive data in a database linked to data defining the movements of the participants.

75. A method of archiving virtual meeting data, comprising:

a storage step of storing data defining a three-dimensional model of the participants in the meeting;

receiving data defining movements of the participants;

generating archive data defining where each person is looking; and

storing the archive data in a database linked to data defining the three-dimensional model amended in accordance with the data defining the movements.

76. A method according to claim 74 or claim 75, wherein, in the receiving step, data is received defining how to amend data stored in the storage step to represent movements of three-dimensional models.

77. A storage device storing computer-useable instructions for causing a programmable processing

apparatus to become configured as an apparatus as set out in any of claims 1 to 36.

78. A storage device storing computer-useable instructions for causing a programmable processing apparatus to become operable to perform a method as set out in any of claims 37 to 76.

79. A signal conveying computer-useable instructions for causing a programmable processing apparatus to become configured as an apparatus as set out in any of claims 1 to 36.

80. A signal conveying computer-useable instructions for causing a programmable processing apparatus to become operable to perform a method as set out in any of claims 37 to 76.

CHARTER MARK

The Patent Office
PATENTS • DESIGNS • COPYRIGHT • TRADE MARKS

INVESTOR IN PEOPLE

122

Application No:	GB 9905191.4	**Examiner:**	Frank D. Moeschler
Claims searched:	1-13, 19-28, 34-68, 74-80	**Date of search:**	30 August 2000

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.R): G4A (AUDB); H4F (FAAX, FDX, FGXD, FGXX, FRG, FRX) H4T (TBEC, TBEX)

Int Cl (Ed.7): G06T-15/00

Other: Online: WPI; JAPIO; EPODOC; INSPEC

Documents considered to be relevant:

Category	Identity of document and relevant passage		Relevant to claims
A	WO 99/65223 A2	(THERMO INFORMATION SOLUTIONS) See pages 6-9, 14	
A	US 5819286	(YANG et al) See whole document	
Y	US 5499146	(DONAHUE et al) See whole document	1, 19, 34, 37, 57, 74 and 75 at least
Y	http://garuda.imag.fr/MPEG4/syssite/syspub/version1/index.html		1, 19, 34, 37, 57, 74 and 75 at least

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

This Page Blank (uspto)